In presenting the dissertation as a partial fulfillment of the requirements for an advanced degree from the Georgia Institute of Technology, I agree that the Library of the Institute shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish from, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate Division when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

7/25/68

EVALUATING AND SCHEDULING PROTOTYPE REQUIREMENTS FOR SUITABILITY TESTING

A THESIS

Presented To

The Faculty of the Division of Graduate

Studies and Research

By

Richard Benway Cole

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Operations Research

Georgia Institute of Technology

June 1972

EVALUATING AND SCHEDULING PROTOTYPE REQUIREMENTS FOR SUITABILITY TESTING

Approved:

Chairman: Dr. Norman R. Baker

Dr. ~~Jerry Connolloy~~

Dr. R. G. Parker

Date approved by Chairman: May 23, 1972

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES vi

LIST OF ILLUSTRATIONS vii

SUMMARY viii

Chapter

I. INTRODUCTION 1

Purpose
Background
Discussion of the Problem
Concept

II. LITERATURE SURVEY 11

Introduction
Current Army Literature in Suitability Testing
Value Measurement
Scheduling
Summary

III. DETERMINATION OF MODEL PARAMETERS 37

Introduction
Background
Identification of Factors
Weighting of Factors
Categorization of Factors
Weighting of Categories
Categorization of Requirements
Summary

IV. DETERMINATION OF MEASUREMENT FUNCTION 71

Introduction
Model Description
Selection of the Model
Application of the Model
Summary

TABLE OF CONTENTS (Continued)

Chapter Page

V. SCHEDULING 87

Introduction
Assumptions
Aspects of Scheduling
Unconstrained Scheduling
Constrained Scheduling
Sub-Test Design
Summary

VI. CONCLUSIONS AND RECOMMENDATIONS 101

General Comments
Conclusions
Recommendations

APPENDICES

A. INSTRUCTIONS TO USERS 104

B. PROCEDURES 112

LITERATURE CITED 115

ACKNOWLEDGEMENTS

I appreciate the time and guidance given to me by Dr. Norman R. Baker, Associate Professor in the School of Industrial and Systems Engineering, during the conceptualization, application, and presentation of this Research. His assistance was essential throughout this research. I also appreciate the assistance and constructive comments given by the other members of my reading committee, Dr. Terry Connolly and Dr. R. G. Parker.

Colonel Charles B. Thomas and other members of the United States Army Infantry Board provided time, assistance, and facilities without which this research would not have been possible. Their cooperation is deeply appreciated. Special thanks are due to Major Alan B. Phillips, the Chief of Methodology and Instrumentation at the Test Board. The assistance, guidance, and encouragement given by Major Phillips were absolutely essential to this research. I also appreciate Major Phillips' acting as an honorary member of my reading committee.

The fine typing job done by Mrs. Peggy Weldon is appreciated.

Finally, I thank my wife, Barbara, for typing the drafts from my scribbled notes and for providing encouragement necessary for the completion of this project.

Special permission was received from the Graduate Division for some variation from thesis manual requirements to fit the needs of this thesis.

LIST OF TABLES

Table		Page
1.	Definitions	41
2.	Number of Personnel Assigning Rank Positions to Each Factor, Test A	49
3.	Ratings of Factors for Test A	49
4.	Number of Personnel Assigning Rank Positions to Each Factor, Test B	51
5.	Rankings and Weights Developed for Factors	51
6.	Number of Personnel Assigning Rank Positions to Categories, Test A	57
7.	Number of Personnel Assigning Rank Positions to Categories, Test B	58
8.	Rankings of Categories, Tests A and B	62
9.	Rankings and Ratings of Categories, Tests A and B	65
10.	Results of Comparisons of Requirements, Test A	77
11.	Rankings of Models Based Upon Comparisons, Test A	77
12.	Ranking of Models Based Upon Simulation No. 1, Test A	79
13.	Ranking of Models Based Upon Simulation No. 2, Test A	79

LIST OF ILLUSTRATIONS

Illustration Page

1. Schematic of Concept 10
2. Schematic of Category Weighting 63

SUMMARY

This research addressed the problem of developing a schedule for the suitability testing of the prototype of a complex item. A sequential approach to the problem is developed. This approach involves ordering the requirements against which the prototype is to be evaluated and then using the ordered set of requirements as a basis for sequencing the tests included in the suitability test.

Emphasis is placed on developing the ordered set of requirements. A model based upon the criteria recommended by Fishburn and by Moore and Baker is developed for mapping the requirements from a randomly arranged set to an unconstrained test sequence. In developing this model, the methods of ranking, rating, successive ratings, and the Delphi technique are studied. These methods are used to assign quantitative measures to qualitatively described requirements and factors for consideration. A linear model is developed which is determined to be an acceptable normative model.

Three methods for evaluating the models are presented. Two of these methods are applicable for discriminating between predictive models. The third method, involving simulation, is found to be a satisfactory method for discriminating between normative models.

The bulk of this research is carried out at a test board of the United States Army. Two actual suitability test are studied.

CHAPTER I

INTRODUCTION

Purpose

The purpose of this research is to develop a method useful in developing a schedule of a suitability test for the prototype of a complex item. A suitability test is a test designed to evaluate the prototype in order to determine if the item represented by the prototype is suitable for production. The overall test of the prototype will usually involve evaluating the prototype against many, often unrelated, requirements. Generally, a separate test is required for the evaluation of one or more requirements. Consequently, the suitability test will actually consist of a series of individual tests, or sub-tests. The specific problem addressed in this research is to determine a method of scheduling the sub-tests, which comprise the overall suitability test, so as to maximize the rate in which information, relative to the potential suitability of the item, is generated during the suitability test.

Background

In the development of a complex item of equipment, it is common for the equipment to undergo a Research and Development (R & D) cycle of several years in length and to incur R & D costs of several million dollars. One of the last phases of the R & D cycle is the development and test of a prototype of the item. The actual test of the prototype can be quite expensive and time-consuming and can directly affect the final cost and final availability date of the end item. Consequently, if this phase of the cycle could most efficiently serve its purpose, then an important portion of the cost and developmental time of the end item could be minimized.

The purpose of the suitability test of a prototype is to provide information upon which a decision of item disposition can be made. The decision usually will be to determine whether the item represented by the prototype should be accepted and placed into production, accepted contingent upon certain modifications, retained for further development, or rejected from further consideration. This decision may be made prior to the completion of the suitability test.

The overall suitability test will consist of a series of intermediate tests each designed to evaluate the prototype against one or more specifications or operational requirements. For example, the item represented by the prototype may be required to have a cruising range of 200 miles, to have the ability to negotiate a 60 degree slope, and to weigh no more than one ton. Then the suitability test may include an intermediate test on operational characteristics designed to evaluate the prototype against the specifications of cruising range and climbing ability and another intermediate test of physical characteristics during which the item would be weighed. In this thesis, the specifications imposed on the item, and to be evaluated during prototype testing, will be referred to as requirements and the intermediate tests designed to evaluate the prototype against one or more operational requirements will be referred to as sub-tests each designed to generate specific information about the prototype. The information accumulated from these sub-tests then serves as a basis for the decision relating to the final disposition of the item.

The time required to make the decision on equipment disposition directly relates to the time required to accumulate sufficient information upon which the decision can be based. Consequently, it is desirable that the sub-tests be scheduled so as to maximize the rate of information generated. This is obviously a particularly important criterion in the scheduling of sub-tests of prototypes of items required for an immediate need. On the other hand, care must be exercised so as to prevent a premature decision on item disposition. Obviously, an incorrect decision could result in accepting an expensive but unsatisfactory piece of equipment, or it could result in delaying the production of a suitable item. Consequently, this research will be directed towards developing a method which will be useful in scheduling the sub-tests so as to maximize the rate of information generated during the test.

The problem of developing a test schedule which will maximize the rate of information generated is compounded and made more important by the fact that there is frequently no predetermined stopping rule upon which the decision on item disposition can be made. For example, it is frequently undesirable to decide before the test that if a certain percent of the operational requirements are not met, then testing will stop and the item will be rejected. This type of stopping rule may be unsatisfactory since the performance of the prototype against other requirements may be so outstanding as to overshadow its failures, or the degree of failure may be more important than the failure itself. For example, the hypothetical item mentioned earlier may have failed both operational characteristics requirements but passed the physical characteristic requirement. An arbitrary stopping rule may have caused the item to be rejected on its climbing ability being limited to 58 degrees and its cruising range being limited to 190 miles, but its amazing weight of only one-half ton combined with its other characteristics make it a very desirable item. Consequently, the decision of item disposition must be based upon the quantitative data of the number of requirements passed and also upon the subjective evaluation of the prototype's overall performance.

The problem of developing a test schedule, with the objective of maximizing the rate in which information is generated and in which no definitive stopping rule can be established, is not specifically addressed in the current literature. However, this is a pressing problem facing R & D organization in general, and the United States Army in particular. The environment studied in the research of this problem will be suitability testing of prototype items within the United States Army.

Discussion of the Problem

The mission of the United States Army Test and Evaluation Command (USATECOM) is to conduct suitability testing of items of equipment which are developed for possible introduction into the Army's inventory. By the time the equipment is submitted as a prototype for suitability testing, it will have completed a Research and Development cycle of from one to twenty years and will have incurred developmental costs of from a few hundred to several million dollars. The items comprise a heterogenous set. For example, some items recently tested were rain parkas, anti-tank missles, underwear, all-terrain vehicles, personnel parachutes, and night vision devices. The characteristics that all of the items for test have in common are that each must be evaluated against several requirements, that there are no definitive stopping rules applicable to these tests, and that it is desirable to maximize the rate in which information is generated during the suitability tests.

Since there are many types of equipment being tested, USATECOM created Branch related Test Boards to test that equipment which is most related to a particular Branch. For example, the Infantry Board at Fort Benning, Georgia is currently testing the Dragon Anti-Tank Missile System whereas the Field Artillery Board at Fort Sill, Oklahoma, tested the new lightweight 105mm towed howitzer recently adopted by the Army. In addition to the Branch Boards, an Arctic Test Center is located in Alaska, a Tropic Test Center is located at the Panama Canal Zone, and a Desert Test Center is located in Arizona. In forming these Test Agencies it was felt that the personnel making the determination of equipment suitability should be familiar with the demands to be placed on the equipment and the environment in which the equipment will be utilized. This is a logical assumption and is based on recognition of the fact that there will be at least some degree of subjectivity involved in making the determination of suitability.

Before a Test Board receives item(s) of equipment for suitability testing, the Board receives literature describing the equipment and the requirements against which the equipment is to be tested. The members of the Board consider the equipment, the need(s) which it is intended to meet, and the requirements against which it is to be tested. If, in their opinion they feel that the requirements are either overly stringent or insufficient for making a determination of suitability, they will recommend changes as appropriate. After the requirements have been firmly established, a test officer is assigned to plan and conduct the actual suitability test.

Regardless of the type of equipment being tested and regardless of which Board carries out the test, an objective of the test is to obtain as much information relative to equipment suitability as rapidly as possible. A schedule of sub-tests designed to meet this objective may or may not be one which provides for the entire test to be completed as rapidly and as economically as possible. With this objective in mind, the test planner must develop a schedule which sequences the sub-tests of the prototype against specific requirements in a manner which will maximize the rate in which information will be generated.

Concept

It is hypothesized that there are certain factors relating to a suitability test which influence the desired sequencing of its sub-tests. These factors must, of course, relate to the amount of potential information which could be gained from executing the sub-test. An illustrative factor pertaining to the amount of information is the importance of the requirements tested. For example, the information gained from evaluating the prototype against an essential requirement would contribute more information upon which to base the decision of item disposition than would evaluating the prototype against a relatively minor requirement. An example of an essential requirement is the requirement that the item be safe to operate since an item that is unsafe to operate could not be considered suitable regardless of its other characteristics. Item color could be a relatively minor requirement since color will frequently not affect the ability of the item to serve its purpose or else it would most likely be relatively easy to change the color. Consequently, it would generally be desirable to evaluate the prototype against the essential requirement, safety, prior to evaluating the prototype against the minor requirement, color.

However, there may be several factors which warrant consideration. In the tests considered in this research, five factors were identified as influencing the desired relative placement of the sub-tests in the testing sequence, and these factors were found to be of varying degrees of relative importance. For example, the importance of the requirement and the potential destructiveness of the sub-test needed for the evaluation of the prototype against the requirement were two factors which were identified. However, it was also determined that the importance of the requirement warranted more consideration, or influence, when deciding where to place the requirement in the testing sequence. Consequently, the factor, Importance, was considered to be more important than the factor, Destructiveness.

In addition to the factors and their relative importance, the degree to which each factor would apply to each requirement must be considered. For example, one requirement may be considered to be critical to item suitability, another requirement may be important to item suitability but not absolutely essential, and a third requirement may be a "nice to have" attribute of the prototype. Consequently, the degree in which the factor, Importance, applies to each of these requirements varies. The possible degrees of applicability of a factor to the requirements included in the suitability test are defined as the categories of the factor.

In this research, methods are developed for identifying, weighting as to relative importance, and categorizing each factor applicable to the suitability test of a prototype. These methods are developed and described in Chapter III.

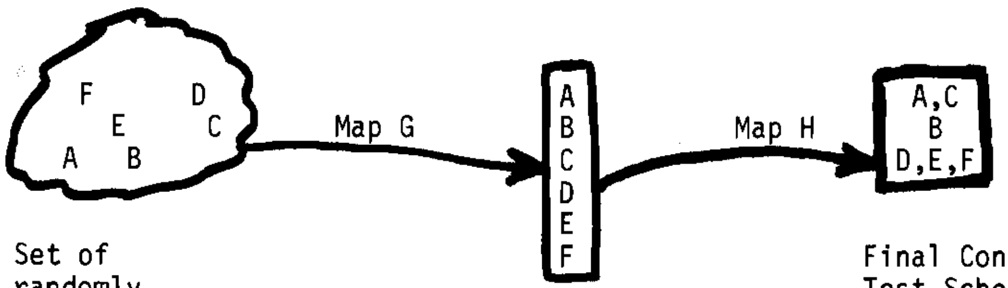
The factors, factor weights, and factor categories are considered to be suitability test dependent. For example, the factors identified as being applicable to the suitability test of one prototype may not be applicable to the suitability test for another prototype; however, the factor weights established for a particular suitability test are considered to be applicable to each of the sub-tests contained in the suitability test. Consequently, identifying the factors, factor weights, and factor categories is done while considering the particular prototype and its required suitability test.

After considering the suitability test as a whole, the next step in developing the desired test schedule is to consider each requirement and its required sub-test. This step involves identifying the category of each factor applicable to each requirement. A method used for this step is also developed and described in Chapter III.

As a result of the analysis of the suitability test and the analysis of each requirement, each requirement will be described in terms of the degree to which the factors apply to the requirement and to the sub-test needed for the evaluation of the prototype against the requirement. For example, consider two hypothetical requirements, requirement A and requirement B. Suppose requirement A is considered to be essential but the sub-test needed is potentially destructive to the prototype, while requirement B is not considered to be essential and the sub-test needed is not potentially destructive. There is obviously a trade-off to be made between the desire to place essential requirements early in the test sequence and the desire to maintain the prototype in a testable condition. These trade-offs become unmanageable when several factors of several categories each must be considered. Consequently a model is developed for mapping the requirements from a randomly ordered collection of requirements to an ordered set of requirements. This model maps the requirement against which the prototype should first be evaluated to the first place in the ordered set. This model and the methods used in mapping are discussed, developed, and applied in Chapter IV.

Finally, after the requirements are ordered, they are assumed to be in the proper sequences for testing. However, in developing the actual test schedule, there may be constraints on testing which require that several requirements be grouped into one sub-test, or which prevent the tests being sequenced as desired, or which affect the test schedule in other ways. Consequently, a second model is then needed to map the requirements from their positions in the ordered set to their final position in the test schedule. This final mapping is discussed in Chapter V.

The concept upon which this research is based, the two stage mapping of requirements from a set of randomly placed requirements to the actual test schedule, is schematically shown in Figure 1.



Set of randomly placed requirements

Set of ordered requirements

Unconstrained Test Sequence, S, is S = A, B, C, ...

Final Constrained Test Schedule Sub-Test 1 Consists of requirements A and C.

Figure 1. Schematic of Concept

CHAPTER II

LITERATURE SURVEY

Introduction

The literature survey is directed to three general areas of investigation: the current US Army literature relevant to the sequencing of suitability tests; the literature relevant to the field of value measurement; and the literature relevant to the field of scheduling.

Current Army Literature in Suitability Testing

An item of equipment which is developed for inclusion in the Army's inventory is subjected to a series of tests during its developmental and production lifetime. These tests generally include the following types which are usually performed in the sequence shown:

1. Research and Feasibility Test, conducted during exploratory and advanced development phase;
2. Developmental Suitability Test, conducted during exploratory and advanced development phase;
3. Engineering Design Test, conducted during engineering development phase;
4. Expanded Service Test, conducted during engineering development and sometimes into production phases;
5. Production Validation Test, conducted during the production phase of the item. (1)

The tests included for a particular item of equipment constitute the item's Coordinated Test Program (CTP) which is the overall or master test plan for the item. A general description of these tests and of the Army's overall testing philosophy for developmental and acquisition testing can be found in (1).

The target of this research is the Expanded Service Test (EST), a type of suitability test. "The EST is a development test (AR 70-10) whcih ascertains that the development of an item is completed and produces the technical and operational test data upon which type classification and initial production decisions may be made." (2).

Included in the EST are tests to determine the degree with which the item meets specified performance standards, to evaluate the training and maintenance test package, and to provide data on the item's overall effectiveness or military worth. Consideration is also given to verify doctrine of use, of organization and tactics applicable, of issue and of logistics and training. The results (both objective and subjective) of the EST, and the Engineering Test immediately preceding the EST, are the basis for the decision (made at Department of the Army Headquarters) of item disposition (1). If the item is accepted for production, it will either go into initial limited production or into major production, depending upon the results of the EST and of the urgency of need for the item. The Expanded Service Test is therefore one of the most important tests in the R & D cycle (2).

The United States Army Test and Evaluation Command (USATECOM) is responsible for the conduct of the EST. USATECOM has directed that in conducting the EST, "The USATECOM objective is to obtain the maximum amount of information for making the determination of suitability in the minimum span of time." (3). USATECOM has further directed that test schedules will be designed to insure efficient programing and utilization of funds and to insure that every effort will be made to minimize the overall test time (4).

However, little concrete guidance is given on how the test should be planned and scheduled to insure that the maximum information will be generated as rapidly as possible. A method is provided in (3) for defining so-called "High Risk" sub-tests and requirements. However, it is shown in (4) that a high risk sub-test may require placement in the latter stages of the test schedule even though "High risk sub-tests will be ... scheduled as early as practical in the test cycle." (4). The purpose of considering the high risk sub-tests early is to address as soon as possible those requirements which will substantially affect the time required to make a determination of equipment suitability.

In discussing (3) with its co-authors, (5,6), it was determined that the regulation is oriented towards identifying major areas of risk but leaves the freedom of test design and schedule to the test officer (5). In other words, the problem of designing a test schedule is not directly addressed but one important aspect of test scheduling is highlighted. Consequently, the test planner derives little practical benefit from the procedures outlined in the publication (7).

An interesting analytic approach to the design of test schedules is presented in (8). However, this approach is based on the assumption that a decision rule exists for classifying the item of equipment. Unfortunately, no such decision rule currently exists, or if it does, neither the Chief of Methodology and Instrumentation at the Infantry Test Board nor this author were able to ascertain its existence (9).

In summary, the results of the survey of Army literature pertaining to suitability testing indicate that there is no published methodology for the scheduling of the EST. It was determined that the EST should be scheduled to provide for the generation of maximum information as rapidly as possible but the procedures which the test planner should use are not specified. Furthermore, since there exists no method for establishing a definitive decision rule, "Judgements by testers do have a place in the EST." (4).

Value Measurement

The specific problem in this research is to determine an efficient method for developing a test schedule which will result in the maximum rate of information flow during the conduct of the test. For the concept of rate of information flow to have meaning, there must be some relative measure of information value. In other words, there must be some measure(s) of value which can be applied to the data generated during the test which corresponds to the worth of the data and to the amount of information which can be gained from the data. For example, assume that the prototype is a rifle. Reams of data on the precise dimensions of the prototype may contain little information upon which a decision of equipment suitability can be based. On the other hand, one or two test firings may generate very little data, but this data may contain significant information relating to the probable accuracy and range of the prototype. This information will probably have a substantial effect on the determination of equipment suitability and consequently may be valuable information.

If the decision on equipment suitability were to be made prior to the completion of the test, there would be a degree of risk associated with this decision. If the decision were to accept the item, then there is always the chance, or risk, that the information obtained from subsequent sub-tests would show that the decision made was in error. Conversely, if the decision were made to reject the item, future information could indicate that the decision was in error. Consequently, the estimated probability that a requirement will be met is a factor which should influence its relative placement in the test schedule. For example, if during the test it were estimated that the probability that the prototype would fail to meet any of the untested requirements were less than one percent, then a determination of equipment suitability could be made with relatively little risk. Consequently, the estimated probability that a requirement will not be met is seen to be a factor not directly related to the "value" of the requirement, but a factor which will probably influence the desired relative placement of the requirement in the test sequence, and consequently relate to the "value" of the overall sequence.

The problem of factor identification will be more fully discussed in Chapter III. It is sufficient at this stage to note that there evidently are several factors which should be considered in developing the test sequence. Some of the factors are directly related to the potential worth of the information to be gained from testing the prototype against the requirement and are seen as influencing the perceived value of the requirement. Other factors may not directly relate to the potential worth of the information to be gained from testing the prototype against the requirement, but these factors do influence the desired relative position of the requirement in the test sequence and consequently relate to the value of the overall sequence.

The problem of determining the desired relative placement of the requirements in the overall test sequence is seen to be similar to the problem of developing a ranking of competing alternatives in other environments. For example, in an R & D environment, potential projects are frequently ranked based upon such diverse factors as probability of technical success, contribution of the project to the advancement of the state-of-the-art, estimated return on revenues, corporate prestige, absolute value of investment required, and a host of other factors which may be applicable to a particular situation. The purpose of the ranking of R & D projects is to provide a basis for the decision of R & D project selection for funding. Since the suitability test is a facet of the R & D process, and since similarities are seen between the problems of ranking R & D projects and ranking or sequencing test requirements, the field of R & D project selection will be investigated for the purpose of selecting a method to be used in sequencing the test requirements.

Two papers which survey the literature relating to the R & D project selection problem have been published. In these papers, one by Baker and Pound in 1964 (10) and one by Cetron, Martino, and Roepcke in 1967, (11) three basic model categories were identified. These categories are the decision theory, economic analysis, and operations research models. Other articles, such as the ones by Moore and Baker (12, 13), Souder (14), and Pessemier and Baker (15) have provided some updating of the reviews in the Baker and Pound and Cetron et al papers. However, the three basic categories initially identified are still descriptive of the current types of models. Also, it is interesting to note that these papers, written eight and five years ago, are almost invariably referenced in the most recent articles in this field.

Even though the R & D project selection problem has been addressed by many researchers, Baker and Pound (10) found that either there has been little testing and use of the methods or else the results from the application of the models have been guarded secrets of those applying the models. Cetron _et al_ (11) reached essentially the same conclusion and added that there was much room for the refinement and improvement of the existing methods. During the search of the more recent literature, this author found the conclusions by Baker and Pound and by Cetron _et al_ still to be valid. Of course, there has been some work published on the application of these models, such as the Moore and Baker article (13) and the Goodwin article (16), but considering the scope of the subject, there still remain relatively few articles on the application and refinement of the theoretic models. One of the underlying motives of this research is to make a contribution in this area.

The economic models such as Risk Analysis (17), and others reviewed by Hurter (18), are not considered applicable to the research problem. The principle reason is that the requirements under consideration here require subjective qualitative analysis rather than the quantitative analysis required for the economic models.

The operations research models, such as the one proposed by Naslund (19) and the one proposed by Hespos and Strassmann (20) are also not considered applicable to this research. The principle reason for the rejection of this approach is the amount and precision of the data required and the computational difficulty of the methods. The data required must be quantifiably specified, such as the specified values of resources required by Hespos and Strassmann or the distribution of event probabilities required by Naslund. In the problem considered in this research, these values are not readily definable. For example, definitive units of value, such as dollars, are not readily applied to the value of the information which could be gained from testing the prototype against a particular requirement. Consequently, a method for assigning some measure of value, probably dimensionless, to information is an objective of this research and a model which presupposes these measures could hardly be appropriate.

Baker and Pound (10) concluded that the decision theory model is applicable to the problem of R & D project selections. The existence of several factors to be considered in sequencing appears to support a similar conclusion for the current problem. The literature search will therefore be confined to models of the decision theory class.

The purpose of this investigation is to determine a method for assigning quantitative measures, which shall be called measures of criticality, to each of the qualitatively described requirements. The requirements will then be sequenced in order of decreasing value of criticality. If such a model could be developed for sequencing the requirements, and particularly if some measure of relative value of the requirements could be related to the assigned measures of criticality, then a significant contribution to the field of suitability testing would be realized (6).

The problem at hand lends itself to the question: "What is the relative value, or measure of criticality, of the requirements if we assume that each of them is competing for early placement in the test schedule?" Therefore, in establishing the sequence, each position in the sequence will be considered in turn starting with the first position. For the first position there will be N possible alternatives, where N is the number of requirements. For the second position there will be N-1 alternatives and so forth. An over-view of the general field of value measurement will be addressed. The primary purpose in addressing the field of value measurement will be to determine a theoretic basis for selecting from among the decision theory R & D selection models. Also, the development of a general theoretically sound model will be attempted so that the model could be applied to more general problem areas outside of the field of suitability testing.

In discussing a unified theory of value, Fishburn (21) offers several considerations.

> As many people view it, decision theory is concerned with the selection of an alternative from a set of alternatives. A particular theory of decision either tries to describe the pertinent factors that are relevant in selection and the way these factors do, in fact, operate in the decision making or selection process, or else it identifies the factors that ought to be accounted for in the selection and states how these factors should operate in guiding the decision maker to his choice (21).

A theory that describes the currently used factors and the way they do operate in the decision process is descriptive or predictive in nature, while identifying the factors which ought to be considered and the way these factors should operate is prescriptive or normative in nature. Excellent studies of predictive theory are discussed by Hurter and Rubenstein (22) and by Einhorn (23, 24).

In this research, emphasis will be placed on developing a normative model. The primary reason for this is that in the testing envirnoment in which this research was conducted, ranking of the factors was not explicitly done during the planning phase of test scheduling. Consequently, there was no benchmark against which the ranking generated by the model could be evaluated. Also, of primary interest to the test personnel involved, was the test schedule itself - not an intermediate ranking of requirements. Consequently, the purpose of the model is to assist the test planner in developing his test schedule, not simply to rank the requirements as he would have done without the model. For these reasons, the relative desirability of the rankings generated by the models considered will not be a factor for explicit evaluation. The relative desirability of the rankings will be implicitly evaluated by evaluation of the test schedules resulting from the rankings and from the degree to which an ordered set of requirements assists the test planner in developing his test schedule.

Fishburn (21) describes value as referring to either preferences among alternatives or to the measures of utility assigned to these preferences. An example of simple preference among alternatives can be found in ranking alternatives in a vector such as $(X_1, X_2, X_3 \ldots)$ where X_1 is preferred to X_2 is preferred to X_3 etc. In this ranking X_1 is described as being more valuable than X_2 is more valuable than X_3 etc. If a measure of utility, $U_i(X_i)$, could be assigned to each alternative, then each alternative could be rated by its measure of utility. An advantage to this would be that quantitative measures of value could be assigned to conflicting alternatives. A model which will rank the alternative and which will assign measures of utility to the alternatives is the target of this research.

In considering the potential usefulness of a theory of value measurement, Fishburn describes a desirable and useful trait to be that which helps the engineer or decision maker in making better decisions in the time available. However, he also reports only qualified success to date due to:

1. Inadequately formulated decision problems;

2. Level of theoretical complexity not readily assimilated by the potential user;

3. Lack of proper instruction by the theory's sponsor on the user's application of the theory;

4. Poorly designed format and program for measuring values. (21)

In considering the desirable characteristics of a unified theory of value, Fishburn lists twelve traits. Of course attempting to develop a unified theory of value which meets all characteristics is certainly beyond the scope of this research, but these characteristics are listed to emphasize what should be considered.

1. Presenting an integrated picture of an individual's preferences, useful for guiding the individual to better decisions;

2. Resolving uncertainties that may attend an individual's belief about his preferences;

3. Presenting an integrated picture of group preferences when the group acts as a decision making unit, useful for guiding the group to better decisions with a clear description of the relationship between group preference and individual preferences;

4. Predicting the actions of "other" individuals and groups of individuals;

5. Interrelating the predictive and prescriptive facets of decision making in a useful and consistent way;

6. Translating rough conceptions of worth and rough statements of objections into precise value terms;

7. Describing the interactions of preference among the factors or variables in multivariate alternatives;

8. Presenting a unified picture of the relationships between preferences and uncertain answers to questions of fact;

9. Describing the relationship and disparities between different and interactive value systems;

10. Specifying a unified method for resolving conflicting value systems;

11. Aiding decision makers in determining whether or not to experiment to gain more information and aiding them to say when a decision ought to be made;

12. Account for the change in value over time and the introduction of new values. (21)

Pessimer and Baker (15) defined three basic sets of methods for determining the relative desirability, or relative value, of the members of a set of items. These sets were categorized as Comparative Methods, Scoring Methods, and Benefit Contribution Methods.

The first set considered, the set of Comparative Methods, includes those methods which require that the evaluator compare the items and assign relative measures of value to each of them. Examples of methods in this set are to have a group of "experts" on the items to be evaluated rate the items or to have the group rank the items.

Eckenrode (25) investigated the relative desirability of several comparative methods and found that all of the methods considered yielded substantially the same results. However, the fastest and easiest method was having the "experts" simply rank the factors under consideration and then correlate the rankings to develop relative measures of value for each

Item (25).

However, these two methods were also evaluated by Goodwin (16). Goodwin found that better results were obtained by having the "experts" rate the items on a numerical interval and by averaging the individual ratings. Not mentioned by Goodwin, but an advantage inherent in the rating method, is that both the rankings and the relative value of each item as perceived by each "expert" is made available to the decision maker(s). Whether this extra data would be desirable in all situations is not known, but it may be a consideration.

Other methods of this same set are successive ratings, successive comparisons, and the dollar metric. These three methods were described and evaluated by Pessimer and Baker (15). Their research indicates that each method is good in that it is generally feasible and yielded results considered to be acceptable. The dollar metric method was determined to be generally better in the environment in which their research was set. However, the method calls for relating a quantitatively defined unit (the dollar) to the items, which in essence calls for considering three entities when comparing two items i.e., the two items and the dollar. In this research, the test requirements do not lend themselves to dollar valuation. Consequently, the dollar metric technique in this environment requires the "expert" to compare "apples and oranges." Also, the dollar metric technique requires making $\frac{n(n-1)}{2}$ comparisons in evaluating n items.

There are other methods included in this first set of so-called comparative methods, but the ones cited are representative. Also, the ones mentioned were the only methods of this type which appeared to be potentially useful in this research.

Most of the techniques included in the Benefit Contribution Methods are the economic models mentioned earlier, relevance trees, and assessment trees . For the reasons previously cited, the economic models will not be considered further. The relevance trees and assessment trees are also rejected as being too complicated for the problem under consideration.

An interesting Benefit Contribution model to determine measures of value for research tasks has been proposed by Nutt (27). This model includes consideration of the value of the task to the overall Research and Development effort of the organization and it also includes consideration of the value of the task to the technical goals of the local unit of the organization. It is interesting that each of these value considerations is computed by multiplying scores assigned to the various contributing factors while the overall task score is computed by adding these two measures of value. Another interesting feature of this model is that it incorporates basically quantitative factors, such as number of technical goals supported, with qualitative factors such as relative importance of the future systems (27). The model is a good example of the application of the concept of determining the value of a project based upon the evaluation of the many factors which contribute to its overall value.

The scoring methods are used to relate quantitative or qualitative considerations to develop a measure of relative value, a score, for each member of a set of items. Basically, the technique involves developing dimensionless and quantified measures of value, or scores, for each consideration, or factor, and arithmetically combining the scores applicable to each item under consideration. This results in a dimensionless score for each item which is taken to be the relative value of the item.

Perhaps the best articles written to date on the scoring model are two papers by Moore and Baker (12,13). In these papers the authors point out that one of the primary advantages of the scoring model is that it allows the explicit inclusion of subjective or qualitative factors for consideration. Perhaps the most important attribute of the scoring model is its inherent ability to generate information (12). However, before the scoring model is lauded as being the panacea for test planners and other decision makers, it must be noted that Moore and Baker found that it is nearly impossible to prescribe how a model should be designed and verified for use in a specific environment. One of the currently unanswered questions in scoring model design is the proper arithmetic form to be used.

One of the best known examples of the application of the scoring model concept is the Mottley-Newton model (26). This model was developed for use in selecting projects for industrial research. In this model five factors each with three categories are specified against which each of the competing projects is evaluated and scored. The five factor scores assigned to each project are then multiplied in order to determine the net score for each project. Of course, the project scores are essentially ratings of each project relative to each other project. Four characteristics of the Mottley Newton model which are subject to question are that the factors against which the projects are to be evaluated are predetermined instead of being an outgrowth of the particular circumstances; these factors each receive equal importance; these factors are each divided into only three categories; and the factor scores are multiplied in computing the overall project score.

A recent example of the scoring model concept is offered by Goodwin (16). In determining the scores to be assigned to major factors, such as cost or performance, Goodwin found that the best method was to have "experts" rate each factor instead of simply ranking them, and then to average the ratings (16). In computing the project's overall measure of value, each project was evaluated against each factor to determine value measures for each project relative to each factor. Each of the value measures was then multiplied by the score determined for the appropriate factor to determine what we shall call the project-factor score. Several models were tried for combining the project-factor scores to derive an overall project score and a simple additive model was found to be superior (16).

In several articles Einhorn (23,24) investigated linear and nonlinear functions in attempting to identify the type of function which best approximates man's response to multivariate stimuli. Einhorn found that for certain decisions the nonlinear, or configural, function produces results more descriptive of the actual mental process of the respondent, than did linear functions. However, Einhorn also found that "...the magnitude of the differences between the fit for the linear and configural models was not large. In addition, even the most configural judges could be fairly well estimated by a linear model."(23) In another article by Einhorn it was determined that nonlinear models more closely correlated to the actual decision processes in about half of the cases where significant differences between the linear and configural models were found, but that on the average there was little difference obtained from the use of linear or nonlinear models (24).

It was mentioned earlier that Moore and Baker found it almost impossible to prescribe how a model should be developed for a specific environment. Results of this literature survey bear them out. However, Moore and Baker do offer guidelines for model construction. Some of their recommendations are summarized as:

1. Environmental considerations should be evaluated in determining the factors for inclusion in the model;

2. There must be some method for measuring the performance of a project relative to each factor;

3. There must be a method for weighting the importance of each factor;

4. A benchmark must be selected against which the results of the scoring model can be evaluated.(12)

Moore and Baker also investigated the computational analysis of scoring models. Among the findings of their research is that "The additive form of the scoring model produced better correlational results..." with other models considered than did the multiplicative model (13). This finding by Moore and Baker in 1969 is basically the same as the finding by Goodwin in 1972.

Of course not all of the literature on scoring models or on methods of value measurement has been discussed in this brief literature survey. However, from the articles mentioned, it is evident that the scoring model concept is being, or at least has been, used under different environments. It is also evident that there is not yet a firm methodology for scoring model construction that is recognized by all users. However, from the results obtained thus far, and since the scoring model is a "good" model which overtly includes subjective considerations and evaluations, it appears that research in this area and the application of the results of past and current research would be valuable.

Scheduling

The field of scheduling is quite broad and touches on a multitude of disciplines. In this literature survey only those aspects of scheduling theory which appear to have some relevance to the specific subject of this thesis will be considered.

In this research, the problem of information value is considered in the construction of the model used to map the requirements into a test sequence. Through the use of the model, measures of criticality are assigned to each requirement. These measures of criticality will now be assumed to be the potential value of the information which may be obtained from the data generated in testing the prototype against the requirement.

Based on the above assumption, the scheduling problem now becomes the general problem of scheduling activities of known potential value with the objective of maximizing the rate at which value is obtained from the activities. If the time required for each activity were the same as the time required for each of the other activities, then the scheduling problem would be solved. The solution would be to schedule the activities so they were sequenced in decreasing order of value. However, the time required for each activity may not be constant, there may be constraints which affect the feasible sequencing of the activities, and it may be desirable that multiple activities be conducted concurrently. In this research, the activities are the tests of the prototype against the specific requirements. These tests were previously defined as sub-tests. Therefore, the specific problems are: (1) to construct sub-tests by grouping requirements for their simultaneous evaluation during the sub-test; and (2) to sequence the resulting sub-tests.

A combinatoral approach to the problem of grouping was considered. This problem is analogous to the problems of determining the smallest number of individuals required to complete a fixed number of tasks or of determining the minimum time required for a fixed number of individuals to complete a fixed number of tasks. The test requirements would be equivalent to the tasks and the sub-tests would be equivalent to the individuals. Unfortunately, there appear to be no computationally good methods of solving problems of this type. Of course complete enumeration is always possible but may be infeasible. For example consider an unconstrained problem of placing r requirements into n sub-tests, an example of the so-called occupancy problem. Feller (27) has shown that the number of possible ways in which this can be done is

$$\binom{n + r - 1}{r}$$

Assume 50 requirements and five sub-tests. Complete enumeration would require considering over 3,000,000 possible assignments. Testing by almost any suboptimum schedule should be completed, or at least well under way, before all possible schedules could be considered!

The problem could be considered as being a form of the so-called traveling salesman problem. The analogy would stem from letting the requirements be the cities that the salesman must visit, the item of equipment be the salesman, and the time required for testing be the distances that the salesman must travel between cities. However, the number of solution procedures advanced for the solution to this problem in the literature (Linear Programing, Dynamic Programing, Network, and Branch and Bound to mention a few achieving moderate success) attest to the difficulty of the traveling salesman problem. Consequently, there is little reason to believe that more success would be realized in solving the grouping problem if it were formulated as a traveling salesman problem.

The problem under consideration in this research is basically a search problem in that it is desirable to locate as much information about the item of equipment as rapidly as possible. Consequently, several search schemes were investigated.

An approach to the problem of search sequences for information systems has been proposed by Baker (28). The problem addressed by Baker was to determine a search sequence (analogous to a test sequence)which would result in information need satisfaction with a minimum expected price in cost and time. The sequence proposed by Baker was to order the search in decreasing value of P_i/C_iT_i where P_i is the probability of obtaining the required information at site i, and C_i and T_i are the cost and time respectively required to examine source i. Baker found that the search sequence was a function only of cost, time, and probability of success (28). Similarly, it is hypothesized that the test sequence desired should be a function only of time and information value for the sub-tests. However, the problem of grouping several information sources into one search site, which would be analogous to grouping several requirements into one sub-test, was not addressed by Baker.

A similar approach to the problem of establishing a search sequence was proposed by Greenberg (29). Greenberg investigated the problem of testing a failed system so as to minimize the expected time required to locate the cause of failure. Letting T_k be the mean time required to test component k and letting P_k be the expected probability that component k has failed, Greenberg proposed that the components should be sequenced for testing in order of decreasing value of P_k/T_k for all k. In other words, he was trying to minimize the expected time required for information need satisfaction and found the appropriate sequence to be only a function of time and probability of need satisfaction. His problem and findings are seen to be quite similar to those discussed by Baker.

The similarity of the problem under consideration in this research and the multiresource constrained Assembly Line Balancing problem was investigated. A Linear Programming (LP) approach was considered but was discarded, principally because Bowen found the LP approach to be unrealistically involved and impractical (30). The same conclusion was offered by Davis and Heidorn (31). However, Davis and Heidorn applied a network approach to the problem but their algorithm was rejected because it is quite complicated, requiring application of both network theory and dynamic programming, and because it was oriented towards minimum job (test) duration without regard to the rate of value obtained.

The Assembly Line Balancing problem was investigated. Helgeson and Birnie (32) considered the problem of determining the minimum process time for a given number of work stations and the converse of determining the minimum number of work stations for a given process time, when only positional precedence constraints are active. If the sub-tests were considered as work stations, some similarity between the test sequencing problem and this assembly line balancing problem is apparent. However, the technique was rejected since it does not guarantee optimality, requires a number of enumerations of possible sequences which could get very large for only a modest number of requirements, and since the value of the work stations or explicitly, sub-tests, was not considered.

Considering the test requirements as being items for manufacture and the sub-tests as being machines on which the items must be processed, the problem has some similarities to the machine sequencing or job-shop problem. An extension of Branch and Bound applied to a graph-theoretic representation of the problem was investigated in (33). However, the technique was rejected due to the number and complexity of the iterations required to obtain a "good" solution, and due to the fact that the objective of the approach was simply to minimize overall test time regardless of the rate of information value generated. In other words, the items for manufacture were assumed to be of equal value. There is no way readily apparent to modify the procedure discussed in (33) to account for the differences in value of the various requirements.

Let the sub-tests be thought of as being jobs of weighted value, and let the overall suitability test be thought of as being a one machine shop. Then the solution to the problem of establishing an unconstrained test sequence is presented by Conway _et al_ (34) as:

> ...the total weighted flow time ... is minimized by sequencing the jobs so that

$$P_1/U_1 \leq P_2/U_2 \leq \cdots \leq P_n/U_n$$

In the above statement P_i is the process time for job (sub-test) i and U_i is the value assigned to job (sub-test) i. In other words, the optimal sequence is a function only of item value and process time. The similarities between the sequences proposed by Baker, Greenberg, and Conway are striking. However, the problem of grouping the requirements into appropriate sub-tests is still unresolved.

A dynamic programming approach to a similar problem was addressed by Gary (35). Gary considered the problem of selecting intermediate test points during manufacture in evaluating a product resulting from the application of a sequence of potentially unreliable operations. The analogy between problems stems from the fact that an intermediate test would include evaluating the results of all operations conducted since the last test. It is a simple bridge to consider the manufacturing operations to be test requirements and Gary's intermediate tests to be sub-tests. However, the analogy is more apparent than real since the operations requirements are in a fixed order. Also, facility and resource constraints are not considered by Gary, making his problem a special case of the general problem considered in this research. Of the general problem, Gary found that the best known algorithms require an amount of time which may be exponential in the number of tasks, which restricts their usefulness to only very small problems (11). Unfortunately, a problem containing 50 or more requirements would not be a small problem.

The problem under investigation could be considered as being an n/m job-shop problem where there are n job requirements (test requirements) which can be assigned to m machines (sub-tests).

Unfortunately, Conway et al state that for the

> ...case, in which each job must be assigned to an individual machine, no optimal procedure has been offered,...Many proficient people have considered the problem, and all have come away essentially empty-handed (34).

The most promising article relating to the constrained sequencing problem was written by Mankekar and Mitten (36). The problem of sequencing a series of tests subject to given precedence and proximity constraints is addressed. The objective is to construct a test sequence based upon the known cost of each test, the known probability that each test will fail, and the grouping constraints mentioned so that the expected cost of the test sequence will be minimized. It would be a simple matter to substitute value for cost in their algorithm and to convert the procedure into a minimization algorithm. The conversion would also require that time be substituted for probability of failure. Converting the units of measure and converting the minimization algorithm into a maximization algorithm would be relatively trivial. However, the major difference between the problem considered herein and the problem discussed by Mankekar and Mitten, is that forming the proximity constraints (grouping of requirements into sub-tests) is a major obstacle in this research whereas Mankekar and Mitten assume them to be given and to be active constraints to the problem rather than being objectives of the problem.

In summary, the scheduling problem investigated in this research is found to be partially solved in the existing literature. The problem of sequencing the sub-tests in an optimum manner can easily be solved based upon the work by Baker, Greenberg, Conway, or Mitten. However, the problem of grouping the requirements into sub-tests in an optimum manner has defied analytical solution. There are some aspects of this problem which relate to the machine balancing problem and to the job-shop problem, but no solution procedures which would be readily applicable to a problem of this complexity are known.

Summary

The purpose of this Literature Survey was to give a brief overview of the current literature pertaining to the fields of Suitability Testing in the United States Army, Value Measurement, and Scheduling applicable to this research.

In the area of suitability testing, it was determined that it was desirable to schedule tests so as to generate the maximum amount of information as rapidly as possible. It was also determined that the EST, the target of this research, is a very important test in the overall testing program carried out by the Army. However, no firm methodology for determining the "value" of a test or for constructing the test was found. There appears to be a void in the literature on these subjects.

In the area of Value Measurement, it was determined that the field is quite broad. It was also discovered that there exist many fields in which the theory is applied and there exist many techniques for the application of the theory of value measurement. The decision theory approach was found to be most promising for this research. The particular vehicle found to be most applicable was the so-called scoring model which has had some use and about which several articles, both descriptive and analytical, were found. However, there appears to be no firm methodology for the application of this model which should make this research both exploratory and applied.

In the area of Scheduling, it was found that one aspect of the problem addressed in this Thesis has been solved. However, no exact nor workable heuristic approach was found which could be applied to the problem as it is currently defined.

CHAPTER III

DETERMINATION OF MODEL PARAMETERS

Introduction

The "Concept" section of Chapter I outlines a sequential approach to the scheduling of sub-tests within the suitability test of a prototype item. The primary concern is to maximize the flow of information upon which to base the disposition decision. This would imply that the sub-test containing the most important requirement(s) should be placed early in the suitability test. In other words, the more important requirements should be placed near the front of the testing sequence.

However, the suitability test as a whole must be considered when scheduling the sub-tests. For example, if an early sub-test resulted in the destruction of the prototype, then subsequent sub-tests may be infeasible and no additional information could be obtained from the suitability test. Also, the information gained from the testing of the prototype against a relative minor requirement may influence the decision on the procedures to be used in subsequent sub-tests and consequently affect the rate of information flow from the suitability test as a whole. Therefore, the desired placement of the requirement in the suitability test is a function of both the importance of the requirement to the decision on item disposition and the effect that testing the prototype against the requirement may have on the overall rate of information flow. Thus, in developing the schedule of sub-tests, trade-offs may be required between two important aspects. These aspects are, (1) the importance of the requirement to the decision of item disposition, and (2) the impact that testing the prototype against the requirement may have on the overall rate of information flow. Of course, the relative position of each sub-test in the suitability test is determined by the characteristics of the requirement(s) included in the sub-test.

The sequential approach developed in this Thesis is based on the fact that each sub-test can be considered as a subset of one or more requirements. Thus, if a model can be constructed which develops a measure that represents the "Requirement Importance vs Effect on Overall Information Flow" trade-off for each requirement, then these measures can be used in specifying and sequencing sub-tests. Analytically, the sequential approach consists of the following steps which should be carried out for each prototype:

1. Identify the factors, j, ($j = 1,2,....,n$) which are relevant to measuring the "Requirement Importance vs Effect on Overall Information Flow" trade-off.
2. Determine the relative importance of each factor for the suitability test considered. To each factor, assign a factor weight, W_j which reflects its relative importance.
3. For each factor, determine the categories, D_j^k ($k = 1, 2, \ldots m_j$) which can be used to describe the requirements with respect to j.
4. Considering each factor independently, determine the relative importance of its categories. Then assign

a category weight, W_j^k for each k which reflects the importance of D_j^k relative to the other categories of j.

5. Compute the net category weight, N_j^k for each category of each factor. This weight reflects the importance of each category relative to each of the other categories. This weight is computed as $N_j^k = (W_j)(W_j^k)$.
6. Determine the category of each factor which is applicable to each requirement. The weight of category, k, of the factor, j, which is applicable to requirement i is $N_j^k(i)$.
7. Develop a function, G, such that

$$C_i = G\{[N_1^k(i)],[N_2^k(i)],\ldots,[N_j^k(i)], \;\ldots\; ,[N_n^k(i)]\}$$

   where C_i is the measure of "Requirement Importance vs Effect of Overall Information Flow" trade-off for requirement i.
8. Use the C_i's to group the requirements into sub-tests and to schedule the sub-tests.

In most of the current literature on model development and application, the model is considered as an entity. However, it appears that a more logical approach in this research would be to consider first the problem of establishing the parameters which will be used in the model and to then consider the functional operator, G. Consequently, this Chapter will be concerned with steps one through six, the determination of the parameters. Chapter IV will be concerned with step seven, the determination of the function, and Chapter V will be concerned with step eight. The Parameters are listed and defined in Table 1.

Background

The procedures used for establishing the parameters were developed through the analysis of suitability tests of two unrelated items of equipment. The schedule for Test A of Item A and the schedule for Test B of Item B were being developed concurrently with this research by members of the Infantry Test Board at Fort Benning, Georgia. Test A was the first test analyzed and the methods used for Test A were improved before being applied to Test B. This improved procedure is found to be satisfactory in the analysis of Test B and is the recommended procedure.

For Test A, six personnel familiar with Test A were briefed on the procedure outlined in Appendix A and each individual was asked to complete the Questionnaires in Appendix A. These personnel were the Chief Test Officer and his administrative assistant, the operations officer for the test, the Chief of Methodology and Instrumentation (M&I) for the Test Board, the Operation Research officer for the Test Board, and a representative from Combat Developments Command. These personnel comprise Test Group A. For Test B, six personnel, forming Test Group B, are also briefed on the procedure and questioned. These personnel are the Chief of the Test Division of the Test Board, the Chief Test Officer, the Chief of the Field Equipment Testing (FET) Branch and his administrative assistant, the Chief of Human Factors Evaluation, and the Chief of M&I. The only officer participating in the analysis of the requirements of both tests is the Chief of M&I.

Table 1. Definitions

Parameter	Symbol	Definition
Measure of Criticality	C_i	A number which reflects the "Requirement Importance vs Effect on Overall Information Flow" trade-off of requirement i.
Factor	j	A consideration applicable to the suitability test. This consideration is assumed to be common, in varying degrees, to all of the sub-tests and requirements. i.e. 1 ≡ Destructiveness
Category	D_j^k	The degree of applicability of j, e.g., D_1^1 ≡ Destructive D_1^2 ≡ Damaging
Factor Weight	W_j	The importance of j relative to the other factors for a particular suitability test when considering the trade-offs.
Category Weight	W_j^k	The importance of category k of j relative to the other categories of j.
Net Category Weight	N_j^k	The importance of category k of j relative to each of the other categories. $N_j^k = (W_j)(W_j^k)$
Functional Operator	G	The arithmetic procedure used in computing C_i.

A five step procedure is developed and used in the analysis of the two tests. These steps, Identification of Factors, Weighting of Factors, Categorization of Factors, Weighting of Categories, and Categorization of Alternatives, are discussed in the remaining sections of this chapter.

Identification of Factors

The identification of the factors which should be considered in developing a sequence of sub-tests of operational requirements is a critical step. The validity of any model, regardless of its precision, can be no better than the parameters used in the model. If a critical parameter, or factor, is not included then the results generated by the model will necessarily be misleading and inaccurate. On the other hand, if the model includes superfulous parameters, or factors, then the results of the model could be equally inaccurate since the results could be biased by the excess parameters.

The problem of factor identification was discussed with the Chief of M&I at the Infantry Test Board at Fort Benning. Three factors were immediately identified in that they were specified by USATECOM (3). These factors are defined as follows:

1. Probability of Failure: The estimated probability that the requirement will not be met.

2. Impact: The influence that the results of the sub-test containing the requirement will have on the determination of item suitability. Also included is the potential destructiveness of the required sub-test.

3. Consequence: The effect on the test schedule if the requirement is not met.

In conference with members of the Infantry Board at Fort Benning, each of the preceding three factors was evaluated. It was concluded that at Test Board level, the estimated probability of failure could range from a best guess based upon experience with similar items, similar requirements and similar tests, to intuitive hunches based upon an almost complete lack of information and background. For this reason, it was agreed that the estimate of probability of failure, a point estimate, would have more meaning for the test officer if the accuracy of this estimate could be included.

The method of using optimistic, pessimistic, and most likely estimates common in Programmed Evaluation and Review Techniques (PERT) was considered but was discarded as being too involved and as not contributing to the solution of the problem. The rationale for this decision was that there was no reason to believe that three estimates based on ignorance would be any more informative to the test officer than would be one estimate. The technique of considering the distribution of the probabilities of failure for each requirement was also considered. In this technique, an estimate of the expected probability of failure, and of the expected variance of the distribution would be needed. This was also rejected as being more involved than productive.

Possibly as a result of being in a testing environment, the concept of Confidence Level was chosen as a factor for consideration. Confidence Level as used in this context is analogous to the confidence level (values of alpha and beta), attached to test results. The rationale behind considering this factor is that the confidence level assigned to an estimate of probability of failure implies to the test officer the expected accuracy of the estimate. Consequently, this information enables the test officer to distinguish between estimates based upon experience or hard data and estimates based upon indefenceable hunches. For these reasons, it was decided that the estimated probability of failure and the confidence level with which this estimate is made are two factors which should be considered when scheduling a suitability test and when building a model to assist the planner in making the schedule.

The second specified factor, Impact, was then considered. The impact that the results of the sub-test evaluating the equipment against a particular requirement will have on the determination of equipment suitability is obviously a factor to be considered in planning the test schedule and in building the model. This factor was, after very little discussion, included in the list of factors to be considered. However, the concept of combining the impact of the information gained from evaluating the prototype against the requirement, with the potential destructiveness of this evaluation, was discussed in depth. It was decided by the groups that in general, the potential destructiveness of a sub-test was a factor that deserved consideration on its own merits regardless of the impact of the requirement. The rationale behind this decision was that if a requirement relates to a delicate or sensitive component of the equipment, its related sub-test should be conducted early in the test program while a potentially destructive sub-test should be conducted as late as possible. Consequently, the factor, Destructiveness, was added to the list of factors for consideration and destructiveness was deleted from the definition of Impact.

The third specified factor, Consequence, was then considered. It was felt that the consequence, or effect, that a sub-test could have on the entire test schedule was definitely a factor for consideration since a sub-test with a potentially disruptive effect could certainly affect the time and cost involved in making a determination of equipment suitability.

As a result of these evaluations, five major factors for consideration and for inclusion into the model were identified by members of the Test Board. These factors, as mentioned in the preceding discussion are:

1. Probability of Failure: The estimated probability that the prototype will not meet the specified requirement. (Relates to the importance of the requirement).

2. Confidence Level: The estimated accuracy of the estimated probability of failure. (Relates to the importance of the requirement).

3. Impact: The importance of the requirement to the potential suitability of the item. (Relates to the importance of the requirement).

4. Destructiveness: The potential destructiveness of the sub-test required for testing the prototype against the requirement. (Relates to the effect on information flow).

5. Consequence: The effect that the results of the sub-test evaluating the requirement would have on the test schedule if the requirement is not met by the prototype. (Relates to the effect on information flow).

In identifying the above five factors as being appropriate for the tests considered, Tests A and B, the members of the test board agreed that these five factors should not be considered as being an exhaustive list applicable to all suitability tests. It was agreed that there should be a flexible method for selecting the factors appropriate to each test analyzed. Consequently, it is proposed that the first step in analyzing a suitability test is to present the above five factors to personnel responsible for conducting the test and asking them to consider the applicability of each factor. In addition to considering the applicability of the factors listed, it is also essential that the test personnel be given the opportunity to add other applicable factors. This step of considering the proposed factors, deleting those not applicable, and adding other appropriate ones was included in tests A and B. Questionnaire 1 of Appendix A was used in this step.

Formally stated, the first step in the procedure developed is to determine the factors against which the requirements should be analyzed in planning the test sequence. A method which is successfully used in this research and which is recommended is to hold a conference with the personnel familiar with the prototype. In this conference the overall test should be discussed and the factors identified. It is also recommened that the five factors mentioned above be considered since these five factors are found appropriate by two different experienced test groups analyzing two unrelated tests.

Weighting of Factors

In their discussion of scoring models, Moore and Baker (12) stress the importance of assigning weights to the factors in order to insure that the model reflects the priorities of the decision makers. Similarly, in the model being developed, it is essential that weights be determined to reflect the relative importance of the factors.

There are several methods available for determining the relative measures of importance of the factors. These methods include simple rank ordering, rating, and several other methods of comparisions. Eckenrode found that many of these methods yield substantially the same results but that the method of correlated simple rankings was the easiest and fastest method (25). However, Goodwin determined that using weights based on the ratings by multiple judges more accurately reflected the actual factor weights than did weightings derived from simple rank ordering (16).

Based upon the findings by Goodwin and upon the fact that the test planners interviewed desired to know both the relative rankings and the relative weights of the factors as perceived by individual members of the test group, a rating technique was adopted.

After considering several rating schemes, the method of successive ratings (15), was chosen. This method was chosen for the following reasons:

1. It is a simple and fast method;
2. It will allow the decision maker to determine the weights considered appropriate by each judge as well as the overall group weights;
3. It forces each judge to develop ratings which he feels to be consistent;
4. The method is intuitively appealing;
5. Very few examples of the application of the method could be found in the literature and therefore the actual application of the technique could have some contributory results.

The first step in applying the method of successive ratings is to have the members of the groups rank the factors under consideration. The purpose of the ranking is to attempt to obtain agreement among the members of the group as to the rank order of the factors. If this agreement can be obtained, then all members of the group will be considering the factors in the same relation when rating them and the ratings should be more meaningful.

For Test A, a simplified version of the Delphi Technique (37) was used. Each of the six officers in the test group were asked to rank the factors in decreasing order of importance. The six officers proposed six different rankings. Each officer was then given the six different anonymous rankings and asked to reconsider his rankings. The group then discussed the rankings and generated a seventh ranking found to be acceptable to all members. It is interesting that the final group ranking is similar but not identical to the ranking that would have resulted from applying Kendall's (38) rank correlation method to the original rankings. The original rankings by each officer, the correlated ranking and the final group ranking are shown in Table 2. (The following abbreviations apply to all tables: Probability of Failure (P(f), Confidence Level (CL), Impact (Imp), Destructiveness (Dest), and Consequence (Cons)).

After the rank of factors was established, five officers of Test Group A used Questionnaire No. 3 of Appendix A to rate each factor as to its relative importance by the method of successive ratings. After these initial ratings were made, a range of scores was selected which compromised between the range of eighty to one-hundred recommended by the Test Officer and the range of twenty to one-hundred recommended by the M & I Officer. The ratings of all officers were scaled to this compromise range and the scaled ratings were averaged to obtain the factor ratings, or weights, used in the model. These ratings are shown in Table 3.

Table 2. Number of Personnel Assigning Rank Positions to Each Factor Test A

Factor	Rank Positions 1	2	3	4	5
P(f)	3	0	1	2	0
Imp	1	3	1	1	0
Dest	2	1	1	0	2
CL	0	0	2	2	2
Cons	0	2	1	1	2

Table 3. Ratings of Factors for Test A.

Factor	Initial Ratings					Scaled Ratings					Final Rating
P(f)	100	100	100	100	100	100	100	100	100	100	100
Imp	98	80	95	94	90	96	90	94	96	95	94
Dest	95	70	80	78	80	90	85	73	86	91	85
Cons	84	30	75	56	20	68	65	67	71	64	66
CL	80	20	70	41	10	60	60	60	60	60	60

Application of the above technique resulted in factor weights found acceptable by the Test Officer for Test A and by the M&I Officer. However, the wide variations in recommended rankings of the factors as shown in Table 1 and the fact that the correlated ranking was different from the final group ranking, makes the technique used in determining the rankings suspect. Consequently the method was slightly modified for application to Test B.

For Test B, the first step in determining the factor weights was to rank the factors according to their relative importance. The technique differed from that used in Test A in that more iterations of the modified Delphi Technique were used. Each officer of the test group was first asked to order the factors in relative importance. Next, each officer was advised as to the number of personnel in the group who assigned each rank to each factor. Each officer in the test group was then asked to develop a second ordering of factors based upon his own convictions and the rankings by the group as a whole. Each officer was then advised as to the number of personnel in the group who assigned each rank to each factor during the second ordering of the factors. Finally, the group discussed the problem of ranking the factors and developed a final group ordering. The number of personnel assigning the rank positions to each factor is shown in Table 4. The correlated rankings resulting from each iteration and the final group ranking were identical. The final ranking is shown in Table 5.

The procedure used for Test B was well received by Test Group 3 and the members of the group accepted the generated rankings. It is interesting to note the rapid convergence of rankings after two iterations

Table 4. Number of personnel assigning rank position to each factor, Test B

Factor	Rank Positions									
	1st Iteration					2nd Iteration				
	1	2	3	4	5	1	2	3	4	5
P(f)	4	0	1	1	0	6	0	0	0	0
Imp	1	3	1	0	1	0	6	0	0	0
Dest	1	0	0	1	4	0	0	0	1	5
CL	0	1	4	1	0	0	0	5	1	0
Cons	0	2	0	3	1	0	0	1	4	1

Table 5. Rankings and Weights Developed for Factors

Factor	Ranking		Weights	
	Test A	Test B	Test A	Test B
P(f)	1	1	100	100
Imp	2	2	94	88
Dest	3	5	85	30
CL	5	3	60	68
Cons	4	4	66	56

in this example. Consequently, since the group found the technique workable, since the members of the group were forced to consider and reconsider the problem of developing rankings, and since each member of the group felt that the generated ranking was acceptable, the procedure used in Test B for determining the ranking of the factors in order of importance is recommended.

After the ranking of the factors was established, the factors were weighted by the technique used for Test A. The rankings and weights assigned to the factors by Test Groups A and B are shown in Table 5. The differences in both rankings and weights as perceived by the two groups point to the necessity of this step in the procedure since it appears that a particular ranking and weighting of the factors would not be applicable to all tests. Whether these differences are due to the differences between the prototypes tested, due to the differences between the groups, or due to the differences in the procedures used is an unanswered question. However, it appears that the prototype tested is the most important factor since the members of both groups agreed that they could rank and weight the factors only for a particular test and that their rankings and weights could be different for different tests.

Formally stated, the second step in the procedure developed in this chapter is to use the modified Delphi Technique to determine the desired order of the factors and then to use the method of successive ratings to assign weights to the factors. These weights then represent the relative importance which should be given to the factors when planning the test schedule.

Categorization of Factors

After the factors which were considered to be important for inclusion into the model have been selected and weighted, the next task is to categorize each of them. Categorization is analogous to the selection of measures of performance discussed by Moore and Baker (12) and the rating of criteria discussed by Mottley and Newton (26).

The categorization of the factors was originally determined by the author of this thesis, the M&I Officer, and the Test Officer for Test A. The categorization was based on guidance published in (3) practical experience fo the Test Officer, the understanding by the M&I Officer of current test methodology, and this author's perception of the problem. This categorization was subsequently verified in the discussion of the problem by each test group.

In categorizing the estimated probability that a requirement would not be met, it was decided to use the point estimate of this probability. The same technique was also used in categorizing the confidence level with which the estimate of probability of failure is made.

The categories of the factor, Destructiveness, were identified and defined as follows:

Destructive: Testing against the requirement is potentially destructive to the test item;

Damaging: Testing against the requirement is potentially damaging to the test item or to components not under test;

Sensitive: The requirement relates to a component which is delicate and which could be easily damaged during the course of unrelated tests;

Stable: The requirement does not require potentially destructive or damaging testing and does not relate to a delicate component.

The categories for the factor, Impact, were identified and defined as follows:

Critical: Failure to meet the requirement is sufficient for declaring the item to be unsuitable;

Important: Failure to meet the requirement is not in itself sufficient for declaring the item to be unsuitable but the requirement will be given major consideration in making the final determination of suitability.

Desired: The requirement will be given some consideration in making the final determination of suitability;

Minor: The requirement will be given little or no consideration in making the final determination of suitability.

Eight categories for the factor, Consequence, were identified by USATECOM (3). Each of these categories was adopted and a ninth category, Stop Testing, was added. The complete unordered list of these categories is shown. The categories are defined as (if the requirement is not met, the consequence to the test plan may be):

Stop Testing: The test will be stopped for an undeterminable length of time or will be terminated;

Test Delay: There will be a test schedule slippage of from one to five days;

Degrade Test: Testing may continue in a degraded mode while the deficiency is being corrected. There will be no test schedule slippage nor significant effect in the determination of suitability of the item under test;

Overtime Required: Retesting or additional work will be required but there should be no test schedule slippage;

Rescheduling: Testing will continue but rescheduling of subsequent requirements will be required. However, neither rescheduling nor retesting should result in test schedule slippage;

Repeat Test: Testing will continue. The failed requirement will require re-evaluation during other planned tests.

Waive: The requirement will probably be waived due to being overly stringent or beyond the current state of the art. Failing the requirement will have no effect on the test schedule.

Nonessential: The requirement will not affect the determination of suitability and failing the requirement will have no effect on the test schedule.

Formally stated, the third step in the procedure is to determine the appropriate categories of the factors for consideration. A procedure which was found successful in this research is to hold a conference with a test group of knowledgeable members of the test board, have these personnel discuss the test and the factors considered, and have the group identify the appropriate categories. The categories mentioned are recommended for initial consideration by the test group since these categories were found appropriate for two unrelated tests by two test groups.

Weighting of Categories

After the appropriate categories of the factors have been identified, it becomes necessary to order and to weight them. The ordering of the categories of the factors Impact and Probability of Failure were found to be consistent for tests A and B and may be consistent for all tests. For example, the ordering of the categories of Impact were Critical, Important, Desired, and Minor which means that it was considered desirable to evaluate the critical requirements first, the Important requirements second, and so-forth. However, the orderings of the categoriesof the other factors were not found to be so clear cut.

Questionnaire 1 of Appendix A was used to establish the desired ordering of categories of factors. Each member of each test group was asked to order the categories considering each factor independently.

For test A, the individual orderings, or rankings, were correlated by Kendall's method of rank correlation (38) to determine the group rankings. This correlated ranking was later accepted as being satisfactory by the Test Officer and by other members of the Test Group. However, the rankings were not discussed within the group, and other rankings were not explicity considered.

For Test B, a modified Delphi Technique similar to the technique used to order the factors was used and several possible rankings were explicitly considered. The technique was found acceptable for ordering both sets. The criteria for acceptability were that the respondents felt confident in applying it and the results obtained were found to be acceptable. The principle advantages of using the modified Delphi Technique followed by group discussion were that:

1. Group concurrence was obtained.

2. The respondents were given an opportunity to defend and explain their rankings. The discussion was a good method of generating information for the test planner and other decisions maker(s).

3. An apparently better ranking was obtained. For example, for the factor, Consequence, the initial rankings would have resulted in a group ranking, call it rank X. However, two respondents were able to convince the rest of the group that rank X would not be the best rank and consequently a new group ranking, call it rank Y, resulted. The group concurred that rank Y was in fact a better ranking than rank X.

The rankings of the categories and the correlated group ranking for Test A are shown in Table 6.

The rankings of the categories for Test B are shown in Table 7. The results shown in this table are particularly interesting for it is

Table 6. Number of Personnel Assigning Rank Positions to Categories, Test A

Category	Rank Position 1	2	3	4	5	6	7	8	9	Correlated Ranking
P(f)										
High	6									1
Medium		6								2
Low			6							3
CL										
High	5		1							1
Medium		5	1							2
Low	1	1	4							3
Imp										
Critical	6									1
Important		6								2
Desired			6							3
Minor				6						4
Des										
Destructive			1	5						4
Damaging		3	3							3
Sensitive	4	2								1
Stable	2	1	2	1						2
Cons										
Stop	5				1					1
Suspend		5	1							2
Delay	1		3	1	1					3
Degrade		1		2		1	1		1	5
Overtime				1	2	1	2			7
Reschedule			2			2	2			4
Repeat				2	2	2				6
Waive							1	5		8
Nonessential								1	5	9

shown that the final rankings of the categories of three factors, (Confidence Level, Destructiveness, and Consequence) are different from the rankings that would have resulted from simply correlating the individual rankings. Whether the final rankings obtained in the application of this

Table 7. Number of personnel assigning rank positions to categories
Test B

Category	1st Iteration Rank Position 1	2	3	4	5	6	7	8	9	Correlated Ranking
<u>Imp</u>										
Critical	6									1
Important		6								2
Desired			6							3
Minor				6						4
<u>P(f)</u>										
High	6									1
Medium		6								2
Low			6							3
<u>CL</u>										
High	3		3							2.5
Medium		6								1
Low	3		3							2.5
<u>Dest</u>										
Destructive	2			4						4
Damage	1	1	3	1						3
Sensitive	1	4	1							1
Stable	2	1	2	1						2
<u>Cons</u>										
Stop	5					1				1
Suspend		5		1						2
Delay	1		5							3
Degrade		1		2		1		1	1	7
O/Time			1		2	1	1		1	6
Reschedule				2	1	2	1			4
Repeat				1	3		2			5
Waive							2	4		8
Nonessential						1		1	4	9

(Continued)

Table 7. Number of personnel assigning rank positions to categories Test B (Continued)

Category	2nd Iteration Rank Position 1	2	3	4	5	6	7	8	9	Corre-lated Ranking	Final Rank-ing
Imp											
Critical	6									1	
Important		6								2	
Desired			6							3	
Minor				6						4	
P(f)											
High	6									1	
Medium		6								2	
Low			6							3	
CL											
High	3		3							2.5	3
Medium		6								1.0	2
Low	3		3							2.5	1
Dest											
Destructive	2			4						4	4
Damage		3	3							2.5	3
Sensitive		3	3							2.5	2
Stable	4			2						1	1
Cons											
Stop	6									1	1
Suspend		6								2	2
Delay			6							3	3
Degrade				3		1			2	7	9
O/Time				1	1	2	1	1		6	5
Reschedule				1	3	1			1	5	4
Repeat				1	2	1	2			4	8
Waive						1	2	3		8	6
Nonessential							1	2	3	9	7

technique are "better" than the ranking obtained by simply correlating the individual rankings or better than the rankings which would have resulted solely from group discussion is an unanswered question. However, within Group B all members of the group perceived the final ranking to be the best of all considered. Also, the test officer, his immediate supervisor, and the Chief of Test Division all concurred that they considered the information generated by the technique to be helpful in their evaluation of the overall test.

Considering the categories of each factor independently, it would appear that the technique was not needed for ranking the categories of Impact and Probability of Failure. As shown in Table 6, there was no doubt as to the proper rankings. The rankings of the categories of the other factors warrant discussion.

In ranking the categories of Confidence Level, the group initially developed a bimodal ranking in that three of the members of the group agreed on one ranking whereas the other members of the group agreed on an inverse ranking. It is doubtful whether the repeated iterations of ranking would have resolved this conflict of rankings since the second ranking was identical to the first. However, in the discussion that followed, each group was able to explain and defend the rationale behind each set of rankings. During the course of the discussions one of the groups was able to win the other over to its line of reasoning and all personnel agreed on the final ranking.

It is interesting that an initial group ranking was computed by correlating the rankings of the categories of Destructiveness, but that a strong bimodal ranking resulted from the second iteration of individual rankings. This conflict of rankings was resolved during the discussion session, and a unimodal ranking was obtained that was different from the initially computed correlated ranking. During the course of resolving the bimodal rankings for the categories of Confidence Level and Destructiveness, each member of the group presented his rationale. Each arguement was apparently considered solely on its logic regardless of the position and rank of the debator, even though the rank structure within the group varied from full Colonel through Major.

There is seen to be a substantial difference between the correlated rankings of the categories of Consequence and the final rankings of these categories. This difference is particularly interesting since the two correlated rankings are seen to be quite similar. This difference can be explained, at least in part, by the fact that the Chief of Test Division, the highest ranking member of the group, stressed that he wanted the category "Degrade" ranked last. Of course his comments biased the results, but it should be noted that he is the final decision maker and the purpose of a model is to assist the decision maker, not to pre-empt him. The correlated rankings and final rankings of the categories for Tests A and B are shown in Table 8.

Finally, the number of different rankings for the categories of the factor, Consequence, is interesting. This spread of rankings suggests that the categories for that factor are not properly defined. Consequently, it is recommended that those categories should be more carefully considered and defined in future applications of the model. In applying the model in this research, the specified categorization was used since it had been suggested in (3).

Table 8. Rankings of Categories, Tests A and B

Category	Test A Correlated Rank	Test B 1st Correlated Rank	Test B 2nd Correlated Rank	Final Rank
Imp				
Critical	1	1	1	1
Important	2	2	2	2
Desired	3	3	3	3
Minor	4	4	4	4
P(f)				
High	1	1	1	1
Medium	2	2	2	2
Low	3	3	3	3
CL				
High	1	2.5	2.5	3
Medium	2	1	1	2
Low	3	2.5	2.5	1
Des				
Destructive	4	4	4	4
Damaging	3	3	2.5	3
Sensitive	1	1	2.5	2
Stable	2	2	1	1
Cons				
Stop	1	1	1	1
Suspend	2	2	2	2
Delay	3	3	3	3
Degrade	5	7	7	9
Overtime	7	6	6	5
Reschedule	4	4	5	4
Repeat	6	5	4	8
Waive	8	8	8	6
Nonessential	9	9	9	7

Whether the differences in rankings shown in Table 8 are due to the differences in the tests under consideration or due to the differences between the groups interviewed remains an unanswered question. However, the fact that there are differences points to the necessity of this step in the overall procedure.

The rationale for assigning weights to the categories in addition to simply ranking them was similar to the rationale for assigning weights to the factors. For example, there is no reason to assume that there is a linear relation among the categories of the factor Impact. If the relation were linear, then simple ranking would suffice. However, assume that the test planner considered that for the test in question it was very important to evaluate first the requirements which are categorized as critical and that Impact would be a relatively minor consideration in sequencing the sub-tests of the other requirements. In this case he would prefer to have the category, Critical, heavily weighted, and slight differences between the weights assigned to Minor and Important.

A weighting based on simple ranking would not provide the appropriate weights in this example as shown schematically in Figure 2.

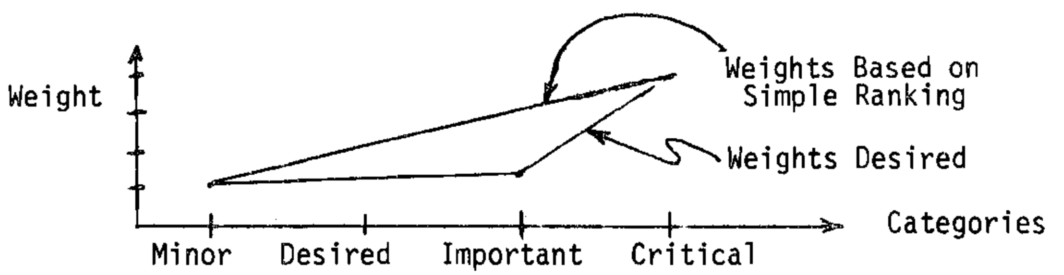


Figure 2. Schematic of Category Weighting

Of course, as mentioned earlier, weights could be computed from the individual rankings using the methods described by Eckenrode (25). However, based upon the findings by Goodwin (16) and the fact that the test planners preferred to know both the weights and the rankings as perceived by the individual members of the test groups, a rating scheme was utilized. Questionnaire No. 2 of Appendix A was used in weighting the categories.

The members of each test group were asked to place the categories of each factor on a scale of one to ten with the categories ranked first receiving the rating of ten and the categories ranked last receiving the rating of one. The other categories were placed on the scale as the respondents saw fit with the only requirement being that the ratings assigned should be rank order consistent with the rankings. The group ratings were computed by simply averaging the individual ratings and scaling the ratings on to an interval of from one to ten. It is interesting that the group scaled ratings in Test A were rank order consistent with the computed group rankings for Test A even though the respondents based their ratings on their individual rankings before knowing what the group rankings were to be. The group scaled ratings for Test B were rank order consistent with the group rankings which is not surprising since the respondents did not rate the categories until after the group rankings were developed. The net results of the two steps of ranking and rating the categories were to obtain a scaled rating of the categories of each factor. The rating, or weight, assigned to each category is then proportional to the desired placement in the test sequence of a requirement falling into the category. For example, assume that the categories of Critical, Important, and Minor received weights of ten, five and four respectively. Then this would indicate that the test planner should emphasize sequencing the critical requirements prior to the Important and Minor requirements. Also, this would indicate that the relative placement of the Important and Minor requirements would deserve little consideration. The rankings and ratings of the categories for both tests are shown in Table 9.

Table 9. Rankings and Ratings of Categories, Tests A and B

Category	Test A Rank	Test A Rating	Test B Rank	Test B Rating
Imp				
Critical	1	10.0	1	10.0
Important	2	8.0	2	8.2
Desired	3	4.2	3	3.7
Minor	4	1.0	4	1.0
P(f)				
High	1	10.0	1	10.0
Medium	2	5.0	2	8.0
Low	3	1.0	3	1.0
CL				
High	1	10.0	3	1.0
Medium	2	5.0	2	6.3
Low	3	1.0	1	10.0
Des				
Destructive	4	1.0	4	1.0
Damaging	3	5.8	3	2.6
Sensitive	1	10.0	2	5.6
Stable	2	6.1	1	10.0
Cons				
Stop	1	10.0	1	10.0
Suspend	2	9.2	2	9.2
Delay	3	8.9	3	8.4
Degrade	5	7.1	9	1.0
Overtime	7	6.0	5	5.3
Reschedule	4	7.2	4	7.0
Repeat	6	6.3	8	2.0
Waive	8	2.4	6	3.5
Nonessential	9	1.0	7	2.8

From Table 9 it is seen that an approximately linear relation among the categories of the factors Probability of Failure and Confidence Level were perceived by the members of Test Group A. However, no other linear relationship between categories is evident. For example, in scaling the categories of Destructiveness, the members of Test Group B perceived little discrimination between the categories Destructive and Damaging but an appreciable difference between the categories Stable and Sensitive. On the other hand, members of Test Group A perceived little difference between the categories Damaging and Stable but significant difference between the categories Damaging and Destructive.

Again, whether the differences between the weights assigned by the different Test Groups are attributable to the differences in tests or to the differences between the groups is an unanswered question. However, the differences in weights assigned by the different groups and the non-linear relations developed within each group point to the desirability of including this step in the overall procedure.

If the factors were considered to be of equal importance, then the task of category weightings would be completed. However, since the factors were not considered equal in importance, it would be incorrect to use the category weights as currently computed. Therefore, each category weight was multiplied by the score assigned to its parent factor in order to derive the net category score for each category. The test planner and the other decision makers for each test were then asked to evaluate the net category scores by any method they chose to determine if the scores were accurately representative of the relative importance of each category in deriving a test schedule. One of the decision makers for test A subjectively recommended changing two of the derived net scores while all other personnel canvassed found the scores to be acceptable for each test.

Formally stated, the step of weighting the categories consists of the following phases;

1. Through the use of the modified Delphi Technique derive the ordering of the categories of each factor applicable to the test.

2. Each member of the test group should then weight the categories by placing the categories of each factor onto a interval of from one to ten. The category ranked first in phase one should be assigned the score of ten, the category that was ranked last should be assigned the score of one, and the other categories placed on the interval as the respondents see fit. However, the ordering of categories on the interval should be the same as the ordering developed in phase 1.

3. Average the scores assigned to each category.

4. Scale the averaged scores onto an interval of one to ten inclusively.

5. Multiply the score developed for each category of each factor by the weight computed previously for the factor.

Categorization of Requirements

The net result of the preceding steps is the determination of the parameters which may be included in the model being developed. The next step is to determine the parameters applicable to each requirement.

This step was carried out with the use of Questionnaire No. 4 of Appendix A. Each member of each group was asked to categorize the requirements by indicating on the questionnaire the Probability of Failure, the Confidence Level, and the Category of each of the other factors applicable to each requirement.

In Test A, these categorizations were evaluated by this author who subjectively correlated the different categorizations. He then developed recommended categorizations which he considered appropriate.

A copy of Questionnaire No. 4 was used to form a composite questionnaire. This composite showed the number of personnel of the Test Group who indicated which category of each of the factors Impact, Destructiveness, and Consequence were applicable to each requirement. Also indicated were the average estimates of Probability of Failure and Confidence Level for each requirement and the categories of the other factors which were recommended. The Test Officer and the M&I Officer were each asked to evaluate the composite questionnaire and to change the recommended categorizations as they saw fit. Both officers considered the recommended categorizations of each requirement to be acceptable.

For Test B, the above procedure was modified. This modification is considered by this author and by the M&I Officer as being an improvement. The primary purpose for the modification is to remove the influence of the model's sponsor from the categorization. Since the members of the Test Group are considered to be the personnel most knowledgeable of the test and of the testing procedures, they should be involved in all of the analysis. Consequently, the influence of a less knowledgeable person, the model's sponsor, on the qualitative analyses is considered to be undesirable.

For Test B, a composite questionnaire was developed similar to the one developed for Test A. However, the composite for Test B did not include recommendations on categorizations. Copies of the composite questionnaires were given to the Test Officer and to his immediate supervisor, the Chief of FET. These officers were asked to evaluate the composites and to recategorize the requirements based upon their own convictions and upon the categorizations by the members of the Test Group shown on the composite. The categorizations by these two officers were similar but not identical for all requirements. The Test Officer was then given a second composite which indicated the original categorizations shown on the first composite and also indicated the second categorizations perceived by the Chief of FET and by the Test Officer. Since the Test Officer has primary responsibility for the Test, he was asked to evaluate the second composite to make the final determination of requirement categorizations. In making the final categorizations, the Test Officer indicated that the procedure used was very helpful in determining categorizations which he considered to be accurate (39).

The procedure to be used in determining the appropriate categories for each requirement is to first require each member of the Test Group to categorize each requirement. Then a composite should be developed which shows the categorizations recommended by the members of the group. This composite is then evaluated by the personnel having primary responsibility for the test. These personnel then make the final determination of categorizations. This step in the procedure is seen to be a simplified form of the Delphi Technique.

Summary

In this chapter the problem of developing the parameters for a model useful in mapping the requirements from an unordered collection of requirements to an ordered set of requirements was considered.

A five step procedure developed and recommended as being applicable to similar type problems in other environments is presented.

1. Identify the appropriate factors through group discussion. Five factors were identified as

being appropriate for the two tests considered. These factors are recommended for consideration, but not necessarily for adoption, for all suitability tests.

2. Rank the factors in order of relative importance by a modified Delphi Technique followed by group discussion. Weight the factors by the technique of successive ratings.

3. Identify the appropriate categories of each factor through group discussion. The categories developed in this Research are recommended for consideration, but not necessarily for adoption, for all suitability tests.

4. Weight the categories by a three step procedure. First, use the modified Delphi Technique to order the categories within each factor. Then weight the categories within each factor by placing them onto an interval of one to ten insuring that the rank order of the categories in the interval is consistent with the rank ordering determined in the first step. Finally, compute the net weight for each category by multiplying the weights determined in the second step by the weights developed for its parent factor.

5. Categorize the requirements using a simplified Delphi Technique.

This procedure resulted in parameters considered to be appropriate by the decision makers of two suitability tests. The procedure was also found feasible in that the personnel interviewed were able to understand it and were able to utilize it. However, the categorization of the factor, Consequence, remains suspect.

CHAPTER IV

DETERMINATION OF MEASUREMENT FUNCTION

Introduction

The problem under consideration in this research is to develop a model which would be useful in sequencing and designing the sub-tests of a suitability test. It is hypothesized in Chapter III that such a model would be of the form

$$C_i = G\{[N_1^k(i)],\ [N_2^k(i)], \ldots,\ [N_j^k(i)], \ldots,\ [N_n^k(i)]\}$$

The parameters of this model, the $N_j^k(i)$, are investigated in Chapter III. The functional operator, G, is the concern of the current Chapter and the design and sequencing of sub-tests is discussed in Chapter V.

In determining G, only linear and simple multiplicative functions are evaluated. Disjunctive and conjunctive functions described by Einhorn (23,24) and logarithmic functions were considered but not included in this research. The primary reason for not investigating these forms is that the potential benefits from the more complicated model would be offset by its computational difficulties. It must be stressed that this research is oriented towards the test planner who cannot be expected to have an operations research or other strong mathematical background. The models evaluated are more fully described in the section titled "Model Description." The procedures used in evaluating and selecting the recommended functional operator and the set of parameters which should be included in the model are described in the section titled "Selection of the Model."

The procedures developed in Chapter III and the model developed and evaluated in this chapter are then applied in developing a test schedule. The application step of this research is discussed in the section titled "Application of the Model."

Model Description

Six models are evaluated in this research. Of primary interest are models Nos. 1, 2,5 and 6. It is hypothesized that either Model No. 1 or model No. 2 is the desired model.

Models No. 1 and No. 2 are respectively formulated as

$$C_i = \sum_{j=1}^{n} N_j^k(i)$$

$$C_i = \prod_{j=1}^{n} N_j^k(i)$$

where n is the number of factors considered.

Models No. 3 and No. 4 are respectively formulated as

$$C_i = \sum_{j=1}^{m} N_j^k(i)$$

$$C_i = \prod_{j=1}^{m} N_j^k(i)$$

where the factor, "Confidence Level," is not included in the set $j = (1,\ldots,m)$.

Models No. 5 and No. 6 are respectively formulated as

$$C_i = \sum_{j=1}^{3} N_j^k(i)$$

$$C_i = \prod_{j=1}^{3} N_j^k(i)$$

where only the three factors specified in (3) are included. These factors are Impact, Probability of Failure, and Consequence.

Consequently, this phase of the research involved designing an experiment in which significant indications of the relative desirability of additive and multiplicative models could be determined. This was the rationale for establishing the set of three multiplicative models and the set of three additive models.

Models No. 1 and No. 2 were included for evaluation since it is hypothesized that one of them is the desired model. Since model No. 6 had been proposed by USATECOM (3) for the purpose of identifying "high risk" requirements, the model was included for evaluation. It should be stressed that the purpose of the USATECOM model as perceived by its sponsors is to identify high-risk requirements whereas the purpose of the model being developed in this research is to map the requirements into their proper positions in a testing sequence.

Models No. 3 and No. 4 were considered in Test A since the factor, Confidence Level, was ranked last in importance by Test Group A and since models No. 3 and No. 4 are essentially compromises between models

No. 1 and No. 2 and models No. 5 and No. 6.

Selection of the Model

Each model is used to compute the measures of criticality for each requirement of Test A. For ease of reading, the term, "Score" is used as being synononomous with the term, "Measure of Criticality."

Based upon the scores computed by the six models, six sequences of requirements are generated. The requirement placed first in each sequence is the one receiving the highest score by the corresponding model. The other requirements are then sequenced in order of decreasing scores. As expected, the sequences are not identical. Therefore, procedures are developed for identifying the most desirable sequence. These procedures are developed for determining a ranking of the sequences and consequently a ranking of the models; for identifying the better function (linear or simple multiplicative); and for identifying whether the model should include the five factors identified earlier or only the three factors specified by USATECOM. It is assumed that since the sequences are determined by the models, the sequence identified as being the most desirable must be the output of the best model. Finally, it is hypothesized that the procedures would be generally applicable for discriminating between similar types of models developed in other applications.

The six sequences associated with Test A were presented to the Test Officer of Test A and to the M&I Officer for their evaluation and ranking. As expected, this procedure is totally unsuccessful in that the officers are unable to rank the sequences. The reason for the lack of success is that a total of 59 requirements are involved and each requirement is described in terms of five factors. If the personnel interviewed had been able to rank the sequences, then with some justification it could be assumed that there is no need for the model. In other words, why develop a model to do something that the test planner could easily do alone? However, since the officers interviewed found ranking the sequences of requirements impossible, or at least very difficult, then the model does address a need.

The second procedure involved identifying the differences between the sequences and asking the same officers to discriminate between the sequences based upon the differences identified. The first step in this procedure is to identify those requirements which received essentially the same ranking based upon mulitplicative models or based upon additive models. Of these requirements, those which received appreciably different rankings by the multiplicative models versus the additive models are identified as selective requirements. For example, requirement No. 5, a selective requirement, of Test A received rankings of 19, 29 and 29 or an average of 28 based on the additive models, and rankings of 57, 58, and 46 or an average of 53 based upon the multiplicative models.

Those requirements which receive approximately consistent rankings by all models are then identified as benchmark requirements. A benchmark requirement which received an average ranking near the median of the additive and multiplicative rankings of each selective requirement is selected. For example, requirement No. 57 of Test A with rankings of 41, 42, 41, 41, 40, and 44 or an average of 41 is selected as the benchmark requirement for requirement No. 5 with average rankings of 28 and 53. A benchmark requirement is selected for each of the selective requirements.

The above process was repeated for those requirements which received appreciably different rankings which could apparently be attributed to the number of factors included in the model.

Twenty-six of the 59 requirements of Test A were used to construct seventeen sets of one selective and one benchmark requirement per set. Nine of these sets were based upon differences in ranking from the additive versus the multiplicative models, function sets, while the remaining eight sets were based upon differences in ranking from the five-factor versus the three-factor models, factor sets.

The Test Officer for Test A and the Chief of M&I were independently asked to select the requirement in each set which should be tested prior to the other requirement in the set. It was hypothesized that the better model could be identified in this manner if the judges consistently selected the requirement which would have ranked above the other requirement if a particular type model or a particular set of factors were used. For example, requirement No. 5 received an average ranking of 28 in the additive models and an average ranking of 53 in the multiplicative models. If requirement No. 5 were selected as deserving testing prior to requirement No. 57, with an overall average ranking of 41, then it would be infered that the additive model was the better model.

The results of the comparisons within each type set and the total comparisons made were used in an attempt to identify the proper arithmetic procedure to be used, the appropriate factors which should be included, and even to correlate the results to particular models and thereby generate a sequence of models based upon their relative desirability. It was hoped that this procedure would assist in identifying the better predictive model and that a simulation procedure to be discussed later would identify the better normative model. Unfortunately, the results of the procedure used to identify the better predictive model did not lead to any definite conclusions except that possibly neither the additive, multiplicative, five factor, nor three factor models were particularly good predictive models. The results of this procedure are shown in Tables 10 and 11. The Test Officer is identified as Judge A and the M&I Officer is identified as Judge B.

Table 10. Results of Comparisons of Requirements, Test A

	Number of Selections Indicating Superiority of			
Judge	Additive Model	Multiplicative Model	Five Factor Model	Three Factor Model
A	4	5	7	1
B	1	8	4	4
Agreements	0	4	4	1

Table 11. Ranking of Models Based Upon Comparisons, Test A

Requirement Sets	Model 1	2	3	4	5	6	t_b*
Function	5.5	2	4	2	5.5	2	.454
Factor	3	3	3	1	6	5	-.40
Total Comparisons	5	1	3	2	6	4	-.28

*t_b is the coefficient of Rank Correlation described by Kendall (38). A value of t_b of +1 would imply perfect agreement between judges A and B whereas -1 would imply complete disagreement.

Since no obvious conclusions can be drawn from the data shown in Table 10 and since none of the values of t_b shown in Table 11 is statistically significant, none of the models is a particularly good predictive model. Of course, there is some indication that the five factor multiplicative model, No. 2, is the superior predictive model. This is indicated by the number of agreements between the judges in Table 10 and by the ranking based upon total selections shown in Table 11. However, these are only indications and are not statistically significant. Finally, the insignificant results could be due to inconsistencies on the part of the judges, instead of due to the inadequacies of the models, or due to the differences between the judges.

A procedure was developed to attempt to identify the best normative model. This procedure involved simulating the flow of information from a test sequenced according to each model. It was hypothesized that if the judges could identify the simulated tests which they considered to be better scheduled, and if these tests could be ranked in order of desirability, then an ordering of the relative desirability of the models would result. For this procedure a seventh sequence based upon random placement of the requirements was generated. This sequence is identified as Model No. 7.

The test officer for Test A, the Chief of M&I, and one of the co-authors (6) of (3) (Judge C) were asked to rank the simulated test sequences of Test A in relative order of desirability. The results of these rankings are shown in Table 12.

Judge C objected (justifiably so) to the fact that the simulations were based on only one iteration. Since models No. 1, No. 2, No. 5, and

Table 12. Ranking of Models Based Upon Simulation No. 1, Test A

Judge	Model 1	2	3	4	5	6	7	W*
A	3	2	4	1	5	6	7	
B	1	5	2	4	3	6	7	
C	2	6	1	4	3	5	7	
Correlated	1	5	2	3	4	6	7	.686

*W is the Concordance coefficient described by Kendall. The value shown indicates agreement between judges significant at the .05 level.

No. 6 are of particular interest in this research, sequences 1, 2, 5, and 6 were simulated seven times and the results of these simulations were correlated and presented to three judges for comparisons. These judges were the Chief of the Test Division (Judge D), the M&I Officer, and the Test Officer for Test B (Judge E). Note that only one officer was involved in ranking both sets of simulations. The results of the rankings of this second set of simulations are shown in Table 13.

Table 13. Ranking of Models Based upon Simulation No. 2, Test A

Judge	Model 1	2	5	6	W*
B	1	3	2	4	
D	1	3	2	4	
E	1	4	2	3	
Correlated	1	3	2	4	.91

*The value of W shown indicates agreement between judges significant at the 0.05 level.

The judges of simulation No. 1 agreed that each of the models under consideration produced sequences superior to the sequence, No. 7, which was based on random placement of the requirements. It was concluded that the additive models (No. 1, No. 3, and No. 5) with rankings of 1, 2, and 4 respectively were superior to the multiplicative models (No. 2, No. 4, and No. 6) with rankings of 5, 3, and 6 respectively. It was further concluded that the five factor models (No. 1 and No. 2) with rankings of 1 and 5 respectively were superior to the three factor models (No. 5 and No. 6) with rankings of 4 and 6 respectively. These conclusions were reinforced by the results of Simulation No. 2.

Models No. 1 and No. 2 were used to construct sequences of requirements for Test B. Each of these sequences was simulated three times and the results were correlated and presented to the Chief of M&I and to the Test Officer for Test B. They concurred that the sequence generated by model No. 1 produced a test schedule superior to that generated by model No. 2. This sequence was so obviously superior that further opinions were not obtained. The results of this simulation further reinforced the conclusions reached above.

Test A had already been scheduled and testing was ready to begin when this research was begun. Since the constrained test schedule for Item A had previously been developed the results of this research project were not used in the actual scheduling of Test A. Also, the requirements for neither Tests A nor B had been explicitly sequenced or evaluated in any manner similar to that developed in this research. Consequently, there existed no benchmark against which the generated sequences or the developed models could be compared. Therefore, since Model No. 1 had been originally hypothesized as being the best model, and since the results of the simulations strongly supported this hypothesis, the ranking of requirements generated by Model No. 1 was used in developing the test schedule for Test B.

In summary, three procedures were used in attempting to identify the best model. The first procedure involved the test personnel attempting to rank the sequences generated by each model. This procedure was found to be unsatisfactory since it required the judges to consider too many variables [(59 requirements) X (5 factors) X (an average of 7 categories per factor) = more than 2,000 decision variables in comparing two sequences].

The second procedure involved the test personnel discriminating between the sequences indirectly. The technique used was to have the judges compare two requirements in each of 26 sets of requirements. This procedure appears to be feasible, but no statistically significant results were obtained. The advantage of the procedure is that each judge is required to consider only 10 variables [(2 requirements) X (5 categories per requirement) = 10 decisions variables] in making each decision. There is no apparent fault in the procedure so it is concluded either that the results indicate none of the models is a particularly good predictive model or that the judges were not consistent in their evaluations.

The third procedure involved simulating the actual results which would have been experienced if tests had been conducted according to each model. This procedure addressed the normative side of model building and was found to be effective. An apparent reason for the success of this procedure is that the test personnel know what is desired in the results of a test schedule, but had previously given little thought to explicitly considering all the characteristics of each individual requirement when designing the schedule. This conclusion is based upon the results of the procedures used and upon interviews with various test personnel.

Application of the Model

The Test Officer for Test B was given sequence No. 1 of the 58 requirements for Test B as generated by Model No. 1. He was then asked to use this ranking of requirements in any way that he saw fit in scheduling the test.

In scheduling the test, the Test Officer first identified the constraints active for Test B. As it turned out, only technological constraints were required and these dictated that the test consist of four sub-tests of multiple requirements. Three of these sub-tests were required to be conducted sequentially and the fourth sub-test consisted of requirements which required evaluation throughout the entire testing period. The requirements which were required to be placed in each sub-test were identified and grouped within their appropriate sub-tests. The ranking of requirements generated by Model No. 1 were then used to order the requirements within each sub-test to form the final test sequence. The ordering of requirements in each sub-test was rank order consistent with the ordering of the requirements in sequence No. 1. Finally, the time and personnel requirements for the sub-tests were identified and a tentative test schedule which required four personnel and two weeks was established. The Test Officer found the ranking of requirements generated by Model No. 1 to be of appreciable assistance when establishing the order in which the requirements would be addressed within each sub-test. He also considered the resulting test schedule to be "optimum," or as nearly "optimum" as he could determine (39).

Prior to the conduct of this research, a tentative test schedule for Test B had been developed. According to the previously developed schedule, a planning figure of 16 weeks was established for the time required to complete the suitability test. Of course, this planning figure is a pessimistic estimate. A most likely estimate of the time required had not been determined.

Through the application of the methods and model described herein, a test schedule was developed with a most likely estimate of the time required being established at two weeks. The test officer did not wish to establish a new planning figure, or pessimistic estimate, until he had re-evaluated all possible contingencies. However, he was confident that the new planning figure would be no more than four weeks. No claims are made that through the use of the procedures and model developed in this research, a test schedule will be developed which will require less than one-fourth of the time which would otherwise be required. However, it appears that the procedures can result in either a substantial savings in test time or a more accurate estimate of the test time required.

The Test Officer estimated that by existing methods it would have taken him more than a week to construct the sequence of requirements generated by the model. Also, he was not confident that his manually constructed sequence would be similar to sequence No. 1 even though he found the sequence to be, as nearly as he could estimate, the "optimum" sequence of test requirements. On the other hand, it took approximately nine and one-half hours or 32 man hours to gather and validate the data, and to generate and validate the sequence used. It is estimated that if all personnel had been familiar with the techniques employed, it would have taken approximately seven and one-half hours or 27.5 man hours to carry out these steps, and that the time could have been reduced to approximately four and one-half hours or 17 man hours if the computations required had been programmed into a computer.

Finally, this author and the Test Officer discussed the resulting test schedule. Without benefit of the sequence generated by the model or of the categorizations of the requirements established earlier, the test officer was asked to justify his test schedule based only upon the verbal descriptions of the requirements and his knowledge of the overall test. He was able to justify convincingly the relative placement of each requirement and found the schedule to be, somewhat to his own surprise, "optimum" from any point of view. Of course, this determination of optimality is based upon subjective judgement and the validity of the conclusions are only as valid as the judgement of the officer making them. However, this officer is an experienced test officer and could reasonably be considered to be an expert in his field. Until a contrained optimization model is developed which will replace expert judgement in qualitative analysis, the opinions of the experts in the field will have to be used in the determination of optimality.

Summary

Six models are developed and evaluated. The purpose of these evaluations is to identify the better functional operator for the model being developed and to identify the best set of parameters which should be included. For the purpose for which the model is developed, it is concluded that the model should be linear and that it should include all parameters previously identified as being pertinent.

In attempting to select the best of the considered models, interrogation of two key decision makers is attempted. This approach is largely predictive in nature in that, if the technique is successful, it would result in selection of a model which would rank the requirements in a manner similar to the ranking that the decision makers would generate on their own. Since no conclusions can be drawn from the results of this technique, other than possibly that the decision makers are neither linear or multiplicative in their decision analysis processes, simulation is attempted. No error in the interrogation procedure is readily apparent and it is recommended that it be considered for use in discriminating between models developed in other environments.

The simulation approach addresses the normative side of value theory in that the results of the rankings rather than the rankings themselves are considered. This approach is found to be successful in that a ranking of sequences (and consequently a ranking of models) is generated with a significant level of concordance among evaluators. The model determined as the best of those considered is a simple linear model. This model includes all factors considered important by the decision makers involved.

Based on the apparent success of this method of verifying a normative model, it is concluded that the technique would be applicable in discriminating between models developed in other environments.

The test planner interviewed finds the sequence of ranked requirements to be very helpful in constructing a test schedule. It is also found that this sequence is generated much faster and more reliably by the model developed than by any informal technique which the test planner would currently be required to use. It is further found that, even though a ranking of requirements prior to scheduling the test would be desirable, this is not currently being done at the Infantry Board, and as far as is known, at any other Test Board within USATECOM. Consequently, it is concluded that use of the model and methods developed should be considered when developing a test schedule. It is further concluded that a model of this type is potentially useful in developing a ranking of any set of multivariate alternatives.

CHAPTER V

SCHEDULING

Introduction

In the preceding chapters of this thesis the problem of determining relative measures of criticality for the operational requirements is considered. A model is developed for determining such measures. The development and analysis of the model is the primary objective of the current Research. Concern in the present chapter deals with consideration of ways in which these measures may be helpful to the test planner when developing a schedule for a suitability test.

The problem under consideration is to determine a procedure for developing an actual test schedule which will result in the optimum rate of information being generated during the test. Of course, the schedule which would satisfy this criterion would be one in which all operational requirements were evaluated simultaneously. However, this alternative is dismissed as being infeasible.

It would be possible to conduct most suitability tests by evaluating the requirements individually and sequentially. If the time required for evaluating each of the requirements were the same, then it is proposed that the best test schedule would be one sequenced by the model previously developed. However, if the time required for testing the prototype against each requirement varied between requirements, then the schedule would be suspect. For example, assume that the time required for evaluating the first requirement in the sequence were one half of the total estimated time required to conduct the entire suitability test. In this case the test planner could choose between the alternatives of conducting one, or all except one, of the sub-tests during the first half of the overall suitability test. The problem of developing a test schedule, in which the time required to conduct a sub-test is a major consideration, is addressed in the section titled "Unconstrained Scheduling."

Technological, or precedence and proximity, constraints are frequently active in suitability testing. These constraints require that one or more sub-tests precede one or more other sub-tests (precedence constraint) or require that two or more sub-tests be conducted concurrently (proximity constraint). The problem of developing a test schedule, in which both the time required to conduct a sub-test and precedence constraints require consideration, is addressed in the section titled "Constrained Scheduling."

It is hypothesized that the suitability test requiring the maximum amount of time is one which is scheduled such that the requirements would be evaluated individually and sequentially. It is further hypothesized that a test schedule containing one or more sub-tests each designed for the simultaneous evaluation of two or more requirements would be a better test schedule. This schedule would be better in that overall test time would be reduced and the rate in which information is generated would be increased. The problem of grouping requirements into sub-tests is discussed in the section titled "Sub-Test Design."

Assumptions

The discussions in the remaining sections of this chapter are based upon several important assumptions. Hence, these assumptions and the rationale upon which they are founded are discussed in the current section.

It is assumed that the measure of criticality of a requirement is a measure of the contribution to the test objective, obtaining the maximum amount of information in the minimum length of time, which will be obtained from testing the prototype against the requirement. Simply stated, it is assumed that the measure of criticality is the same as the value of the information which will be gained from evaluating the prototype against the requirement. This assumption is based upon several considerations:

1. The factors and categories discussed in Chapter III were identified, ranked, and weighted with the objective of developing a score for each requirement. This score was to be a measure of the relative desirability of placing the requirement first in the overall test schedule.

2. The overall objective which was kept in foremost consideration when developing and evaluating the model previously discussed was to schedule a test to maximize the rate in which information is generated during the test.

3. Sequencing the requirements by their measures of criticality resulted in a test schedule which apparently maximized the rate in which information was generated.

4. From (1) it is assumed that the measure of criticality is based upon measures of value assigned to the factors and categories considered. From (2) it is assumed that the model operates upon several measures of value to develop one measure of value. From (3) it is assumed that there is a direct relationship between the value of information generated from testing each requirement and the measure of criticality assigned to the requirement.

It is assumed that the value of a requirement is the same as the value of the information to be gained from evaluating the prototype against the requirement. This assumption follows from the initial assumption given above.

It is assumed that the value of a sub-test is the sum of the values of the included requirements. This assumption is based upon the fact that the model developed is an additive model. This assumption is also supported by a similar hypothesis proposed by Nutt (27). Explicitly the hypothesis stated that the total value of a project was equal to the sum of the values of the project to two different organizations.

It is assumed that there is a linear relation between the value of the information obtained from a sub-test and the length of time which has been spent on the sub-test. In other words, one-half of the total information which will be contributed through a sub-test will be available after the sub-test has been half completed. This assumption was judged valid by test personnel interviewed.

It is assumed that the time required to complete a sub-test is the same as the time which would be required to test the prototype against the most time consuming requirement included in the sub-test. Other requirements in the sub-test will have no effect on the length of time required to conduct the entire sub-test and evaluation of all requirements within the sub-test will begin and end simultaneously. This assumption was found to be reasonable by test personnel interviewed. This assumption is also based upon the fact that forecasting the time required to test a requirement or the time required to complete a sub-test of many requirements is not an exact science at test board level. Consequently, calculations of precise time forecasts are considered to be a waste of time and effort.

It is assumed that the personnel required to conduct a sub-test is the sum of the personnel which would be required to individually evaluate each of the requirements. Test personnel interviewed indicated that this assumption is frequently valid.

Aspects of Scheduling

In any scheduling environment there are several aspects which the schedule planner should consider. These aspects will include the objective(s) of the schedule and constraints on schedule design. Common objectives are to schedule an activity such that it will be completed as rapidly and/or as economically as possible. Common constraints are limitations on manpower or other variable resources and limitations on facilities or other fixed resources.

In scheduling the EST it has been established that the overall objective is to schedule the test so as to obtain as much information relating to item suitability as rapidly as possible. It is also desirable that the total test time and total test cost be minimized, but these objectives are secondary to the objective of maximizing the rate of information flow early in the testing period.

Some of the common constraints on test scheduling at the Infantry Test Board, and assumed to be common to other Test Boards, include limitations on manpower, facilities, and money; statistical requirements on replications; and technological constraints such as the requirement that users of the equipment must be trained to properly operate it before testing of the equipment can begin (9).

Of course the constraints on testing particular items will vary making a determination of all of the constraints applicable to all EST's difficult or impossible. For example, in Test B mentioned earlier, the only constraint active was technological which dictated that the sub-test evaluating the pre-operational inspection requirements precede the sub-test evaluating training requirements and methods. The bulk of the test, the sub-test evaluating the functional performance of the equipment, was then required to be scheduled last. On the other hand, for Test A, the technological constraints were active as were constraints on the availability of manpower, the availability of required sophisticated facilities, the rate of the availability of test items, and other constraints.

Unconstrained Scheduling

In addition to the assumptions mentioned earlier, it is assumed that there are only two aspects which the test planner must consider. First, the objective of maximizing the rate in which information is generated and second, is the fact that the amount of time required to conduct the sub-tests varies from sub-test to sub-test. It is further assumed that the value of each sub-test and the time required for each sub-test is known. The value of the sub-tests were determined through the use of the model previously developed and the times were determined by other means.

Procedure A, a two step procedure, is recommended for determining the test schedule under these conditions. In Step 1 compute the value of T_i/V_i for each $i \varepsilon \{1,n\}$ where T_i is the time required to conduct sub-test i, V_i is the value of sub-test i, and n is the number of sub-tests.

In Step 2, construct a sequence, S(A), as follows:

$$S(A) = (s_1, s_2, \ldots, s_i, \ldots, s_n : T_1/V_1 \leq T_2/V_2 \leq \cdots \leq T_n/V_n)$$

Sequence S(A) is then the desired sequence of sub-tests upon which the test schedule should be based.

This procedure is a direct application of Theorem 3-10 stated and proved by Conway et al (34). The procedure is diagrammed as Procedure A in Appendix B.

Constrained Scheduling

Assume that the assumptions and aspects previously stated as being applicable for Procedure A are active. Further assume that precedence constraints are active. The problem of developing a testing sequence in this environment is similar to the problem of developing a constrained least cost testing sequence described by Mankekar and Mitten (36).

Procedure B is recommended for the development of a test schedule under these circumstances. Basically, this procedure involves isolating those sub-tests for which the precedence constraints are active and then systematically satisfying the constraints. After this has been done, procedure A is applied in a manner which does not violate any of the constraints previously satisfied. The procedure can easily be carried out by hand for relatively small problems or can easily be coded for large problems. The storage capacity required for such a program would be directly proportional to the number of sub-tests considered. This procedure is diagrammed as Procedure B in Appendix B.

Procedure B

The following computational algorithm is given for utilization of Procedure B.

Step 1: Form two sets of sub-tests. Let Set 1 consist of those sub-tests for which precedence constraints are active, i.e., those sub-tests which must precede one or more other sub-tests and those sub-tests which must be preceded by one or more other sub-tests. Let Set 2 consist of those sub-tests for which no precedence constraints are active, i.e., those sub-tests which may be placed anywhere in the testing sequence. Steps 2 through 15 refer to Set 1 only.

Step 2: Order the sub-tests in Set 1 by Procedure A. Index the sub-tests according to their relative position in S(A) with the first sub-test in the sequence being denoted s_1.

Step 3: Form an M x M matrix $R = \{r_{ij}\}$ where:

$r_{ij} = 1$ and $r_{ji} = 0$ if sub-test i must precede sub-test j;

if $r_{ij} = 1$ and $r_{jk} = 1$, then $r_{ik} = 1$;

$r_{ij} = 0$ otherwise;

M is the number of sub-tests in Set 1.

Step 4: Form a matrix $R' = \{r'_{ij}\}$ identical to matrix R.

Step 5: Set the index k equal to 1.

Step 6: Consider each pair of sub-tests i and j. If $r_{ij} = 1$ and sub-test i does precede sub-test j in the current sequence set $r'_{ij} = 2$.

Step 7: If $r'_{ij} \neq 1$ for all i and j go to step 15.

Step 8: Scan R' to determine if $r'_{ik} = 1$ for any i. If there exists an i such that $r'_{ik} = 1$, go to Step 9. If $r'_{ik} \neq 1$ for all i, set k equal to the index of the next sub-test in the current sequence and repeat this step.

Step 9: Form set T_k of all sub-tests i for which $r_{ik} = 1$.

Step 10: Apply procedure A to the set T_k to form the ordered set T'_k.

Step 11: Place the ordered set T'_k immediately in front of sub-test s_k.

Step 12: Consider each pair of sub-tests i and j for which $r'_{ij} = 2$. If sub-test j now precedes sub-test i, set $r'_{ij} = 1$.

Step 13: Set k equal to the index of the first sub-test in the current sequence of sub-tests in Set 1.

Step 14: Go to Step 6.

Step 15: Label the current sequence S(A)'. $S(A)' = \{s'_i\}$.

Step 16: Apply procedure A to Set 2. Label the resulting sequence S(A)". $S(A)" = \{s''_j\}$

Step 17: Form sequence S(B) from sequences S(A)' and S(A)" by iteratively integrating the s''_j into S(A)' such that $V'_i/T'_i \leq V''_j/T''_j \leq V'_{i+1}/T'_{i+1}$. S(B) is the desired sequence for testing.

<u>Proof of Finiteness of Procedure B</u>

1. For any sub-test s_k, there can be only one set T_k since $s_i \in T_k$ if $r_{ik} = 1$.

2. After T'_k is placed in front of s_k, s_k can never be placed in front of any element of T'_k.

Assume $s_j \varepsilon T_k$ and $r_{kj} = 1$;

Since $s_j \varepsilon T_k$, then $r_{jk} = 1$;

But $r_{jk} = 1 \Longrightarrow r_{kj} = 0$ by definition of R.

Therefore $r_{kj} = 0$;

Therefore $s_k \notin T_j$.

3. From (1) and (2) it is evident that there can be no more than M iterations of the procedure since each sub-test is considered once and only once in satisfying the precedence constraints.

Proof of Optimality of Sequence S(B)

1. Consider sequence $S(B) = (\ldots., s_i, s_{i+1}, \ldots..)$ and sequence S(B)' identical to S(B) except that s_{i+1} precedes s_i.

2. In S(B), s_i precedes s_{i+1} either as a result of steps 9 through 11 or as a result of steps 1 or 17.

3. Assume s_i precedes s_{i+1} as a result of steps 9 through 11.

Then $r_{i\,i+1} = 1 \Longrightarrow s_i$ must precede s_{i+1}

Therefore S(B)' is infeasible.

4. Assume s_i precedes s_{i+1} as a result of steps 1 or 17.

Then $V_i/T_i \leq V_{i+1}/T_{i+1}$

Therefore S(B)' is sub-optimum by Conway's Theorem 3-10.

This completes the proof.

Sub-Test Design

In the two sections immediately preceding this one, it was assumed that the sub-tests were predetermined. In this section the problem of designing a sub-test will be discussed.

It is assumed that the only constraint is the personnel constraint.

That is only N personnel are available for commitment to the sub-test. It is further assumed that V_i, P_i, and T_i are known where these variables are the value, the personnel required, and the time required for requirement i respectively. The problem now becomes one of designing the best set of sub-tests which could then be sequenced by procedure A.

This problem was alluded to in Chapter II and seen to be analogous to the n/m job shop problem for which no efficient analytical solution procedure has been offered in the current literature. Consequently, the following procedure, Procedure C, is presented for consideration. This procedure has received very little testing, has not been applied to an actual suitability test, and is submitted here only as a technique for consideration. However, it is simple, it is intuitively appealing, it could easily be coded for a computer with a relatively small storage capacity, and it can be easily carried out by hand.

Basically, the procedure involves forming four sequences of requirements and alternatively drawing from the sequences to form trial sub-tests. The sub-test which contains the most value is then selected as the best of the trial sub-tests and the procedure is repeated until each requirement is assigned to a sub-test.

The following computational algorithm is given for utilization of Procedure C.

Step 1: Apply Procedure A to the requirements to form Sequence 1. For those requirements in which $T_i/V_i = T_j/V_j$ place the more time consuming requirement first in the sequence.

Step 2: Construct Sequence 2 of requirements with the requirements being ranked in order of increasing value of personnel required. Resolve ties by placing the more time consuming requirement first.

Step 3: Construct Sequence 3 of requirements with the requirements being ranked in order of decreasing value of T_i. Resolve ties by placing the more valuable requirement first.

Step 4: Construct Sequence 4 of requirements with the requirements being ranked in order of decreasing value of V_i. Resolve ties by placing the more time consuming requirement first.

Step 5: Construct a base-sub-test by including in the sub-test the first consecutive requirements from Sequence 3 until the inclusion of the next requirement in the sequence would violate the personnel constraint. Call this base sub-test BT.

Step 6: Construct three tentative sub-tests as follows:

a. Add the first consecutive requirements from Sequence 4 to BT until the inclusion of the next requirement in the sequence would violate the personnel constraint. Next add the first consecutive requirements from sequence 1 to the current sub-test until the inclusion of the next requirement would violate the personnel constraint. Finally, to this sub-test add the first consecutive requirements from Sequence 2 until the inclusion of the next requirement would violate the personnel constraint. Label this sub-test BT_1.

b. Construct sub-test BT_2 in a manner similar to constructing BT_1. However, in forming BT_1 requirements were added to BT from Sequences 4, 1, and 2 in that order. In forming BT_2 add requirements to BT from Sequences 1, 4, and 2 in that order.

c. Construct sub-test BT_3 by adding the first consecutive requirements from Sequence 2 until the inclusion of the next requirement in the sequence would violate the personnel constraint.

Step 7: From sub-tests BT_1, BT_2, and BT_3 select the sub-test with the greatest value. Note that the time required for each sub-test is the same as the time required for each of the other sub-tests since each sub-test is based upon BT. Consequently, this step involves selecting the sub-test with the minimum value of T/V.

Step 8: Delete from sequences 1, 2, 3, and 4 those requirements included in the sub-test selected in Step 7.

Step 9: If each of the requirements have been included in selected sub-tests, go to Step 10. Otherwise return to Step 5.

Step 10. Apply procedure A to the sub-tests generated to form Test Sequence S(A).

Step 11: Scan S(A) until the first sub-test is found in which the personnel constraints are not active. Call this sub-test k with test time required being T_k and personnel required being P_k. If no such sub-tests are located, then go to Step 14.

Step 12: Continue to scan S(A) until the first requirement R_j is found such that $T_j \leq T_k$ and $P_j \leq N - P_k$. Place R_j into sub-test k. If no such requirement is located, then return to Step 11. Scan immediately below sub-test k.

Step 13: Return to Step 10.

Step 14: Stop. The current sequence is the desired sequence which should be the basis of the testing schedule.

There are several characteristics of Procedure C which make it intuitively appealing. First, in forming sub-test BT, the requirements needing the longest test times are grouped together. This step provides for concurrent testing of the longer requirements which tends to minimize overall test time. In alternately considering Sequences 4 and 1, an effort is made to include as much value as possible into the time which is required to be used. Finally, Sequence 2 is always considered last in an effort to place as much value as possible into the sub-test where the personnel constraint is the primary consideration.

This procedure is diagrammed as Procedure C in Appendix B.

Summary

The purpose of this chapter has been to introduce some considerations of which the test planner must be cognizant. These considerations differ from the factors previously discussed in that they primarily reflect the constraints which bear on the design of a test schedule.

Some procedures were proposed which may be useful in the design of a test schedule. The complete development and verification of these procedures, and the development of ones which could be used when multiple constraints are active, are beyond the scope of this research. However, it is proposed that the area of scheduling discussed in this chapter is one in which further research could provide a meaningful contribution.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

General Comments

Conclusions are drawn and recommendations made throughout the body of this thesis. However, the principal conclusions and recommendations will be reiterated in the current Chapter.

The primary objective of this research was to develop a method useful in scheduling the sub-tests of the suitability test of the prototype of a complex item. This objective has been met by the development of a model which maps the requirements against which the prototype must be evaluated from a random collection of requirements to an ordered set of requirements. The development of the model was done in a unique manner by first developing a method for generating the parameters which should be included and then developing the functional operator.

Of secondary importance in the research was the objective of developing a method or procedure for generating test schedules in a constrained environment. This problem is discussed and some procedures are recommended for consideration.

Conclusions

The major conclusions drawn from this research are:

1. The unique sequential method used in developing the model is a desirable technique since both the parameters and the functional operator of the model are developed and verified.

2. The model developed is general in nature and primarily based upon this research, research by Fishburn, and research by Moore and Baker which make it applicable for the analysis of multivariate alternatives in other environments.

3. The model is user-acceptable in that untrained personnel who do not necessarily understand, or accept, the overall concept, are able to understand, accept, and apply each step in the method.

4. Neither the linear nor the simple multiplicative models are found to be good predictive models.

5 The linear model evaluated is found to be a better normative model than is the simple multiplicative model.

6. The methods used for model verification are general in nature and are applicable for discriminating between multivariate alternatives.

7. The technique of obtaining group rankings using a modified Delphi Technique followed by group discussion is found to be more desirable than the technique of Kendall's method of correlating the rankings of the members of the group.

8. The technique of Successive Ratings is found to be user-acceptable and to provide acceptable ratings of subjectively described alternatives.

Recommendations

It is recommended that this research be the basis for further study in the following areas:

1. The methods and models developed should be applied to other suitability tests to further verify and refine them in the specific environment in which they were developed. The model should be refined

to make it dynamic and interactive.

2. The methods and models developed should be applied to similar type problems in other environments to verify their general applicability.

3. A procedure should be developed which would provide guidance for the applicability of the model and methods to a particular suitability test. It is not recommended that the model and methods be arbitrarily applied to all tests since the benefits derived may not justify the "cost" of their application to the testing of relatively unsophisticated or inexpensive items.

4. Research should be done in the suitability testing environment on the development of a predictive model which could be compared to the normative model and methods developed in this research.

5. The scheduling procedures discussed in Chapter V should be refined and applied.

APPENDIX A

1. PURPOSE: The purpose of this paper is to outline the methods to be used in conducting risk analysis of a suitability test.

2. DEFINITIONS: For purposes of this paper the following definitions apply:

a. Probability of Failure: The estimated probability that the item under test will fail to meet a prescribed requirement.

b. Confidence Level: The estimated accuracy of the prediction of probability of failure. The confidence level of a prediction will effect the amount of weight which will be given to the probability of failure. A prediction given with a high confidence level will be given more consideration than will be given a prediction with a low confidence level. The confidence level is also a reflection of the amount of information known about the item under test and about the requirement in question. A low confidence level implies that little is known about the item and/or the requirement while a high confidence level implies prior knowledge. For this reason, the confidence level may be a factor for consideration when conducting a risk analysis and when developing a test plan.

C. Impact: The importance of the prescribed requirement to the determination of suitability of the item under test.

Four categories are given:

Critical - Failure to meet the requirement is sufficient for declaring the item under test to be unsuitable.

Important - Failure to meet the requirement is not in itself sufficient for declaring the item to be unsuitable but the requirement will be given major consideration in making the final determination of suitability.

Desired - The requirement will be given some consideration in making the final determination of suitability.

Minor - The requirement will be given little or no consideration in making the final determination of suitability.

d. Destructiveness: The potential destructiveness of the test required to test the item against the requirement. Consideration is also given to requirements which relate to sensitive or delicate components of the item under test.

Four categoreis are given:

Destructive - Testing against the requirement is potentially destructive to the test item.

Damaging - Testing against the requirement is potentially damaging to the test item or to components not under test.

Sensitive - The requirement relates to a component which is delicate and which could be easily damaged during the course of unrelated tests.

Stable - The requirement does not require potentially destructive or damaging testing and does not relate to a delicate component.

e. Consequence: The effect that the item's failure to meet the requirement would have on the test schedule. See page A-3 of TECOM Reg 70-34.

3. GENERAL:

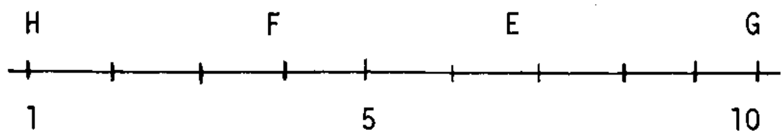
a. The TECOM objective is to obtain the maximum amount of information for making the determination of suitability in the minimum span of time. Additional objectives are specified for a particular suitability test. Therefore, when conducting a risk analysis it is essential to give full consideration to all applicable test objectives and to the general TECOM objective.

b. The purpose of risk analysis is to assist the test planner in developing a test schedule which best meets the applicable test objectives.

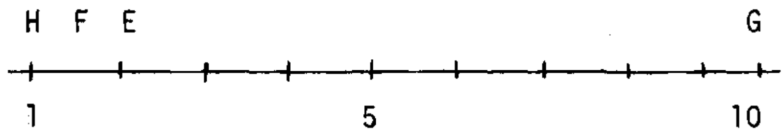
4. PROCEDURE:

a. The first step in risk analysis is to identify the characteristics of requirements which should be given consideration in scheduling a suitability test. In carrying out this step it is important to consider each characteristic independently of specific requirements and of other characteristics. The categories of each characteristic and their influence on scheduling are identified. (Questionnaire No. 1)

b. The second step is to determine the relative importance of each category of each characteristic. Again, each characteristic should be considered independently. For each characteristic, the category which should be tested first will be given a score of 10 and the category which should be tested last will be given a score of 1. Other categories will be scaled appropriately. For example, assume a characteristic is given four categories, E, F, G, and H. Also assume that in step one above, it was determined that the order of testing should be G, E, F, and finally H. If the relative importance of the categories were constant they would be scaled as follows:



If it were very important that category G be tested first but that there was little difference in the other categories, the scaling would be as follows:



See questionnaire No. 2.

c. The third step is to determine the relative importance of each characteristic indentified in steop one. The technique to be used consists of four or more steps.

1. List the characteristics in decreasing order of importance.

2. Give the most important characteristic a score of 100.

3. Score each of the other characteristics relative to the most important characteristic.

4. Repeat the score given to the least important characteristic as determined in step 3. Then score each characteristic relative to the least important characteristic.

5. If the scores in step 3 are not consistent with the scores in step 4, continue these steps until two sets of consistent scores are determined. Consistency after one iteration of steps 3 and 4 would be very unusual but consistency after three to five iterations should be expected.

See questionnaire No. 3

d. The fourth step is to categorize each requirement for the test in question.

See questionnaire No. 4.

QUESTIONNAIRE NO. 1

1. The purpose of this questionnaire is to identify the requirement characteristics which should be considered in conducting this suitability test.

2. Five characteristics have tentatively been identified. Indicate the consideration that should be given to each of these characteristics in developing the test plan. Identify other characteristics which you feel to be important.

3. Considering each of the following characteristics independently, indicate the desired order of testing by circling the appropriate number for each category:

a. Probability of Failure:

Requirements with:	Should be tested:		
A. High Probability of Failure	1st	2d	3d
B. Medium Probability of Failure	1st	2d	3d
C. Low Probability of Failure	1st	2d	3d

b. Confidence Level:

Requirements with:	Should be tested:		
D. High Confidence Level	1st	2d	3d
E. Medium Confidence Level	1st	2d	3d
F. Low Confidence Level	1st	2d	3d

c. Impact:

Requirements which are:	Should be tested:			
G. Critical	1st	2d	3d	4th
H. Important	1st	2d	3d	4th
I. Desired	1st	2d	3d	4th
J. Minor	1st	2d	3d	4th

d. Destructiveness:

Requirements which are:	Should be conducted:			
K. Destructive	1st	2d	3d	4th

L. Damaging	1st	2d	3d	4th
M. Sensitive	1st	2d	3d	4th
N. Stable	1st	2d	3d	4th

e. Consequence:

Requirements which could:	Should be conducted:								
O. Stop Testing*	1st	2d	3d	4th	5th	6th	7th	8th	9th
P. Suspend Testing	1st	2d	3d	4th	5th	6th	7th	8th	9th
Q. Delay Testing	1st	2d	3d	4th	5th	6th	7th	8th	9th
R. Degrade Testing	1st	2d	3d	4th	5th	6th	7th	8th	9th
S. Require Overtime	1st	2d	3d	4th	5th	6th	7th	8th	9th
T. Cause Rescheduling	1st	2d	3d	4th	5th	6th	7th	8th	9th
U. Require Repeat	1st	2d	3d	4th	5th	6th	7th	8th	9th
V. N/A, waiver	1st	2d	3d	4th	5th	6th	7th	8th	9th
W. N/A, nonessential	1st	2d	3d	4th	5th	6th	7th	8th	9th

*This consequence is not listed in 70-34. Stop testing means that the test will be stopped for an undeterminable length of time or terminated.

QUESTIONNAIRE NO. 2

1. The purpose of this questionnaire is to determine the relative importance of each category of each characteristic. For the sake of uniformity each characteristic will be given a spread of 10.

2. Probability of Failure: Three categories of Probability of failure are given, Categories A (.7 to 1.00), B (.4 to .6) and C (0 to .3). Scale the relative importance of these categories by placing the letter codes on the appropriate place of the scale:

1 2 3 4 5 6 7 8 9 10

3. Confidence Level: Three categories of confidence level are given which are similar to the Categories of probability of failure. Scale the relative importance of the categories (D, E, and F):

1 2 3 4 5 6 7 8 9 10

4. Impact: Four categories of impact are given (G, H, I, and J). Scale the relative importance of these categories:

1 2 3 4 5 6 7 8 9 10

5. Destructiveness: Four categories of destructiveness are given (K, L, M, and N). Scale the relative importance of these categories:

1 2 3 4 5 6 7 8 9 10

6. Consequence: Nine categories of consequence are given (O, P, Q, R, S, T, U, V, and W). Scale the relative importance of these categories:

1 2 3 4 5 6 7 8 9 10

QUESTIONNAIRE NO. 3

1. The purpose of this questionnaire is to determine the relative importance of the characteristics identified in Questionnaire No. 1.

2. This questionnaire should be completed as follows:

a. Under "CHARACTERISTIC" list the characteristics in decreasing order of importance.

b. Under the first empty column give the most important characteristic a score of 100. Score each of the other characteristics relative to the most important characteristic.

c. There should now be a score assigned to each characteristic. The score assigned to each characteristic should be greater than or equal to the score assigned to each of the characteristics listed below it. If so, continue; if not, repeat steps a. and b.

d. Under the next column assign to the least important characteristic the score which was assigned to it in step b. Score each of the other characteristics relative to the least important characteristic. If the scores in this and in the immediately preceding columns are identical, stop. Otherwise, continue steps b. and d. until two adjacent columns of identical scores are obtained.

CHARACTERISTIC

I							
II							
III							
IV							
V							
VI							
VII							

QUESTIONNAIRE NO. 4

1. The purpose of this questionnaire is to categorize each Operational Requirement of the test in question.

2. For each OR considered independently:

a. Under P(f) put your estimate of the probability of failure (.0 to 1.0).

b. Under CL put your confidence level in your prediction (.0 to 1.0).

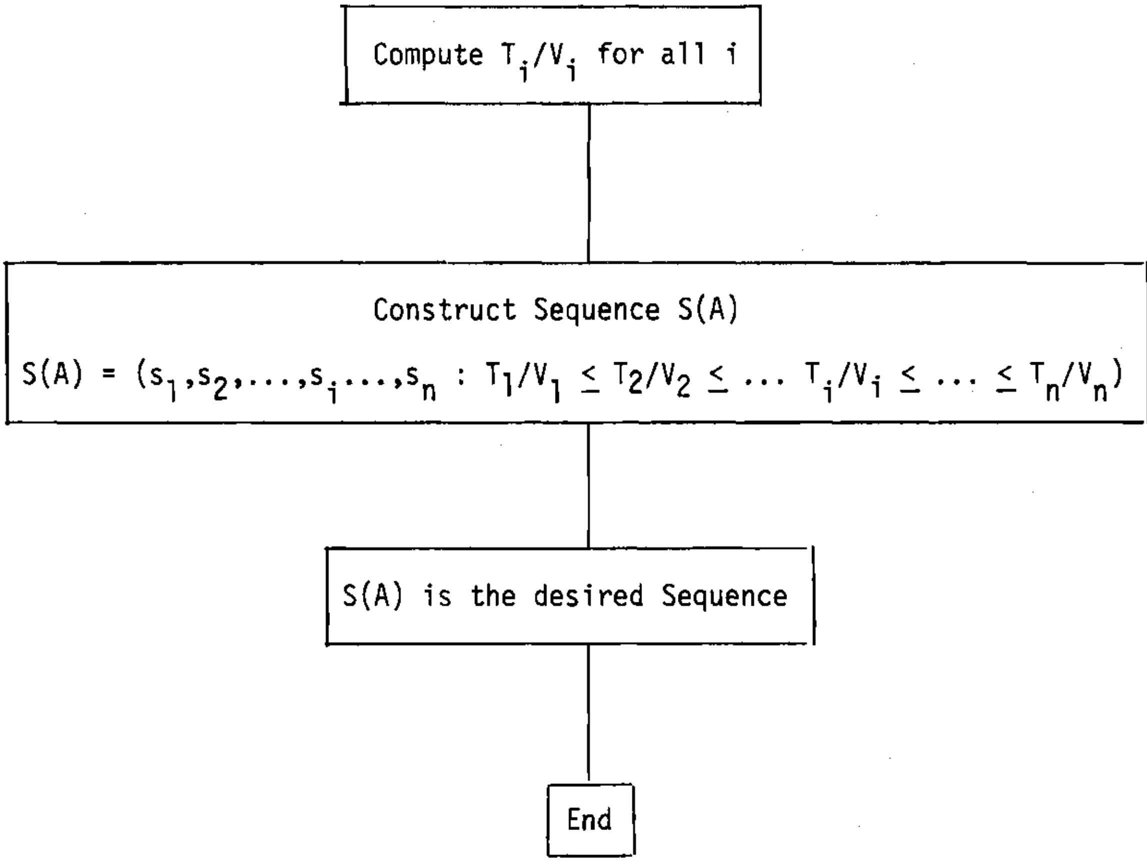
c. Under impact place a mark under the appropriate category for each OR.

d. Repeat step c for destructiveness and for consequence.

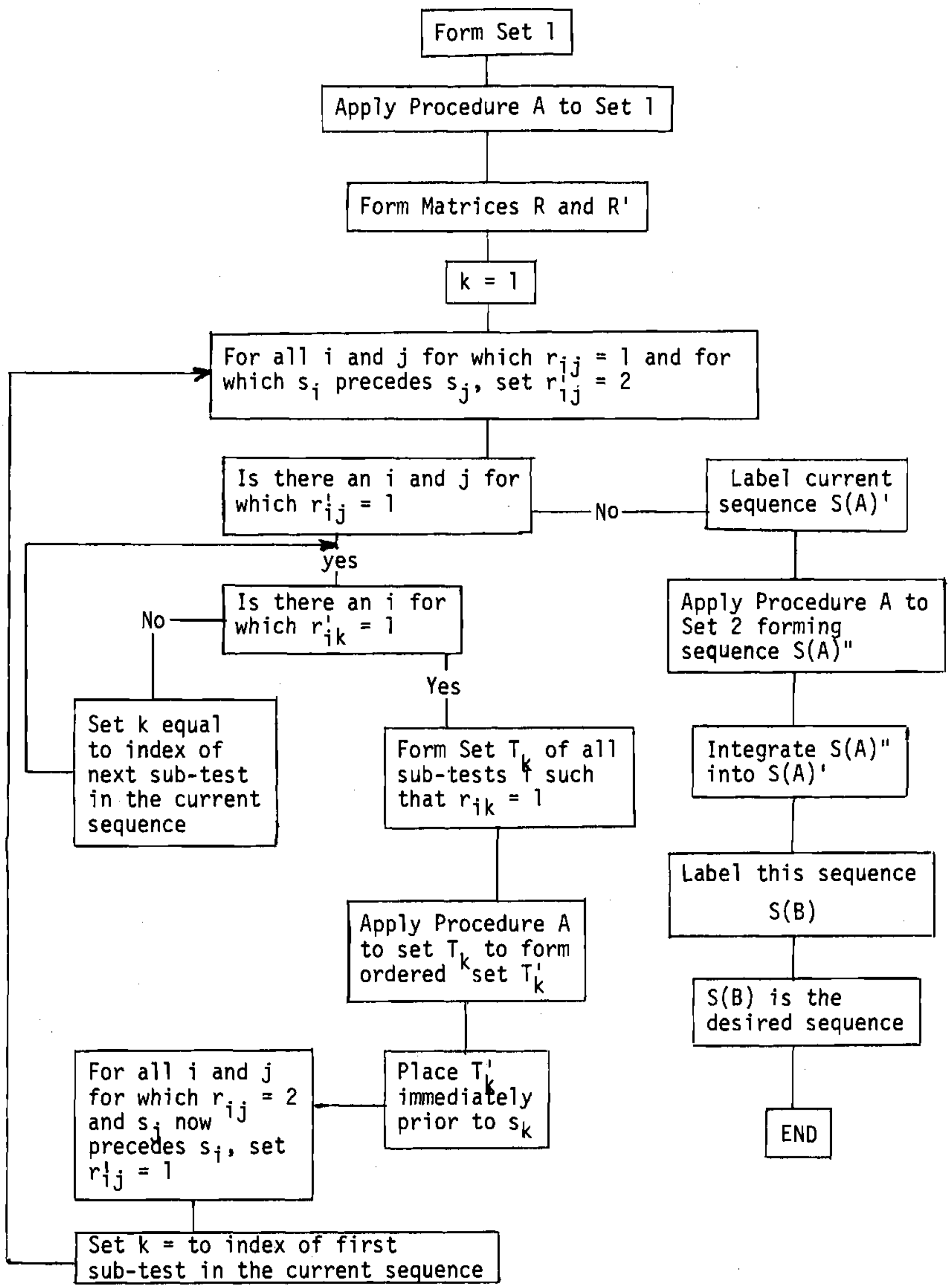
			IMPACT				DESTRUCT				CONSEQUENCE									
OR	P(f)	CL	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	COMMENTS
1																				
2																				
3																				
4																				
5																				
6																				
7																				
8																				
9																				
10																				
11																				
12																				
13																				
14																				
15																				

APPENDIX B

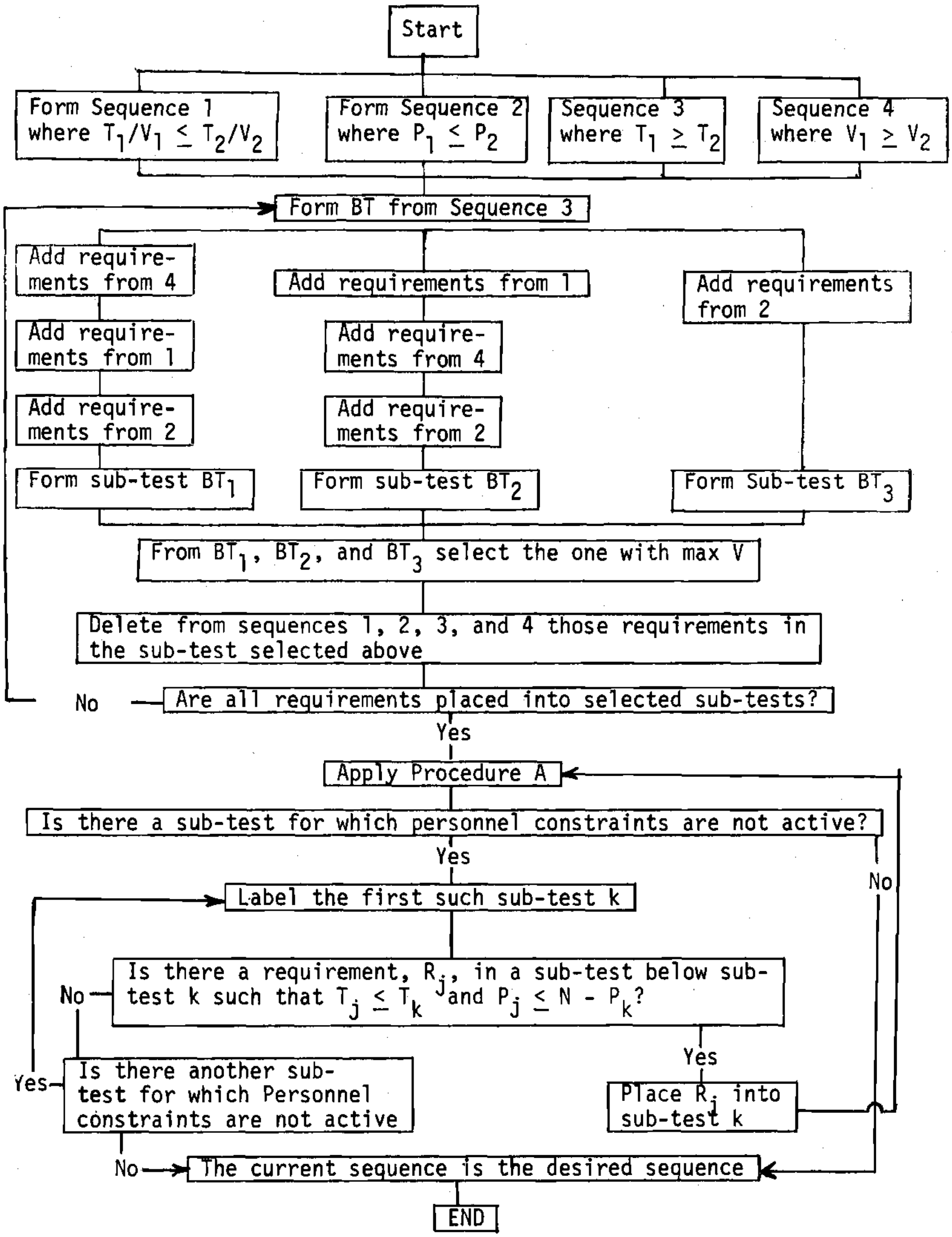
Procedure A



Procedure B


Form Set 1
Apply Procedure A to Set 1
Form Matrices R and R'
k = 1
For all i and j for which r_{ij} = 1 and for which s_i precedes s_j, set r'_{ij} = 2
Is there an i and j for which r'_{ij} = 1
No
Label current sequence S(A)'
yes
Is there an i for which r'_{ik} = 1
No
Yes
Apply Procedure A to Set 2 forming sequence S(A)"
Set k equal to index of next sub-test in the current sequence
Form Set T_k of all sub-tests i such that r_{ik} = 1
Integrate S(A)" into S(A)'
Label this sequence S(B)
Apply Procedure A to set T_k to form ordered set T'_k
S(B) is the desired sequence
For all i and j for which r_{ij} = 2 and s_j now precedes s_i, set r'_{ij} = 1
Place T'_k immediately prior to s_k
END
Set k = to index of first sub-test in the current sequence

Procedure C


Start
Form Sequence 1 where $T_1/V_1 \leq T_2/V_2$
Form Sequence 2 where $P_1 \leq P_2$
Sequence 3 where $T_1 \geq T_2$
Sequence 4 where $V_1 \geq V_2$
Form BT from Sequence 3
Add requirements from 4
Add requirements from 1
Add requirements from 2
Add requirements from 1
Add requirements from 4
Add requirements from 2
Add requirements from 2
Form sub-test BT_1
Form sub-test BT_2
Form Sub-test BT_3
From BT_1, BT_2, and BT_3 select the one with max V
Delete from sequences 1, 2, 3, and 4 those requirements in the sub-test selected above
No
Are all requirements placed into selected sub-tests?
Yes
Apply Procedure A
Is there a sub-test for which personnel constraints are not active?
Yes
No
Label the first such sub-test k
Is there a requirement, R_j, in a sub-test below sub-test k such that $T_j \leq T_k$ and $P_j \leq N - P_k$?
No
Yes
Yes
Is there another sub-test for which Personnel constraints are not active
Place R_j into sub-test k
No
The current sequence is the desired sequence
END

LITERATURE CITED

1. Headquarters, Department of the Army, Test and Evaluation During Development and Acquisition of Material (AR 70 - 10), Washington, D. C., 1971.

2. Headquarters, Department of the Army, "Conduct of Operational Test and Evaluation," (Letter), Washington, D. C., 1971.

3. Headquarters, U. S. Army Test and Evaluation Command, Risk Analysis for Suitability Tests (USATECOM Reg. 70-34), Aberdeen Proving Ground, Maryland, 1971.

4. Headquarters, U. S. Army Test and Evaluation Command, Documenting Test Plans and Reports (USATECOM Reg. 70-24) Aberdeen Proving Ground, Maryland, 1971.

5. LTC. William Echavarria, Interview, Aberdeen Proving Ground, Maryland, March, 1972.

6. George A. Schroeter, Interview, Aberdeen Proving Ground, Maryland, March, 1972.

7. LTC. Santiago A. Garcia, Interview, Fort Benning, Georgia, March 1972.

8. Ralph Shear and Malcolm S. Taylor, "An Information Theoretic Approach to Risk" (TR 10), Aberdeen Proving Ground, Maryland, May, 1971.

9. Maj. Alan B. Phillips, Interview, Fort Benning, Georgia, 1972.

10. N. R. Baker and W. H. Pound. "R and D Project Selection. Where We Stand." IEEE Transactions on Engineering Management. Vol EM VII, No. 4, Dec 1964.

11. Marvin J. Cetron, Joseph Martino, and Lewis Roepcke. "The Selection of R & D Program Content - Survey of Quantitative Methods." IEEE Transactions on Engineering Management, Vol EM-14, No. 1, March 1967.

12. John R. Moore and Norman R. Baker. "An Analytical Approach to Scoring Model Design - Application to Research and Development Project Selection." IEEE Transactions on Engineering Management Vol EM-16, No. 3, Aug. 1969.

13. ----- "Computational Analysis of Scoring Models for R & D Project Selection." Management Science, Vol 16, No. 4, Dec. 1969.

14. W. E. Souder, "A scoring Methodology for Rating Project Selection Models," AIIE Transactions, March 1972.

15. Edgar A. Pessemier and Norman R. Baker, "Project and Program Decisions in Research and Development" R & D Management, Vol 2, No 1, Oct. 1971.

16. Paul G. Goodwin, "A Method for Evaluation of Subsystem Alternate Designs." IEEE Transactions on Engineering Management Vol EM-19 No 1, Feb. 1972.

17. David B. Hertz, "Risk Analysis in Capital Investment." Harvard Business Review Vol 42, No. 1, Jan.-Feb. 1964.

18. Arthur P. Hurter, Jr., "A Selection Review of Investment Analysis, Part 1." Unpublished paper, Northwestern University, Evanston, Illinois.

19. Bertil Naslund, "Mathematical Programing Under Risk." Indian Journal of Operations Vol 1, 1965.

20. Richard F. Hespos and Paul A. Strassmann. "Stochastic Decision Trees for the Analysis of Investment Decisions". Management Science Vol II, No. 10, Aug. 1965.

21. Peter C. Fishburn, "On the Prospects of a Useful Unified Theory of Value Engineering." IEEE Transactions on System Science and Cybernetics Vol SSC-2, No. 1, Aug. 1966.

22. Arthur P. Hurter and Albert H. Ruberstein. "Research Allocation Decision Making Mechanisms in the Private Sector." An unpublished paper. Northwestern University, Evanston, Illinois, Feb. 1969.

23. Hillel J.Einhorn, "The use of Nonlinear, Noncompensatory Models in Decision Making" Psychological Bulletin, Vol 73, No. 3, 1970.

24. -----, "Use of Nonlinear, Noncompensatory Models as a Function of Task and Amount of Information." Organizational Behavior and Human Performance Vol 6, 1971.

25. Robert T. Eckenrode, "Weighting Multiple Criteria." Management Science, Series A, Vol 12, No 3, Nov. 1965.

26. C. H. Mottley and R. D. Newton, "The Selection of Projects for Industrial Research." Operations Research Vol 7, No. 6, November-December 1959.

27. A. B. Nutt, "An Approach to Research and Development Effectiveness" IEEE Transactions on Engineering Management Vol EM-12, No. 3, Sept. 1965.

28. Norman R. Baker, "Optimal User Search Sequences and Implications for Information Systems Operations." American Documation, Vol 20, No. 3, July 1969.

29. Harold Greenberg, "Optimal Test Procedures Under Stress" Operations Research, Vol 12, No. 5, Sept.-Oct. 1964.

30. E. H. Bowen, "Assembly Line Balancing by Linear Programming." Operations Research, Vol 8, No. 3, May-June 1960.

31. Edward W. Davis and George Heidorn. "An Algorithm for Optimal Project Scheduling under Multiple Resource Constraints." Management Science, Vol 17, No. 12, Aug. 1971.

32. W. B. Helgeson and D. P. Birnie, "Assembly Line Balancing using the Ranked Positional Weight Technique." Journal of Industrial Engineering, Vol XII, No. 6, Nov.-Dec. 1961.

33. M. Florian, P. Trepant and G. McMahon, "An Implicit Algorithm for the Machine Sequencing Problem." Management Science, Vol 17, No. 12, Aug. 1971.

34. Richard W. Conway, William L. Maxwell and Louis W. Miller, Theory of Scheduling. Addison-Wesley Publishing Company. Reading, Mass., 1967.

35. M. R. Gary, "Optimal Test Point Selection for Sequential Manufacturing Processes," The Bell System Technical Journal, Vol 51, No. 1, Jan. 1972.

36. P. S. Mankekar and L. G. Mitten, "The Contrained Least Cost Test Sequencing Problem", The Journal of Industrial Engineering, Vol XVI, No. 2, March-April 1965.

37. O. Helmer and N. Rescher, "Epistomolgy of the Exact Sciences," Management Science, Vol. 6, 1959.

38. M. C. Kendall, Rank Correlation Methods. Charles Griffin and Company Limited. 4th Edition, London,1970.

39. Maj. Conrad A. Novak, Interview, Fort Benning, Georgia, March, 1972.